



OPEN ACCESS

EDITED BY

Xiaoqing Ru,
Hebei University of Engineering, China

REVIEWED BY

Wei Liu,
Xiangtan University, China
XianFang Tang,
Wuhan Textile University, China

*CORRESPONDENCE

Zejun Li,
✉ lzjfox@hnit.edu.cn
Mingming Qi,
✉ 20210301@wzut.edu.cn
Lichao Zhang,
✉ zhanglichao@szit.edu.cn

RECEIVED 09 October 2025

REVISED 20 October 2025

ACCEPTED 30 October 2025

PUBLISHED 26 November 2025

CITATION

Wang P, Guo Y, Li Z, Tang D, Qi M, Zhu Z and Zhang L (2025) DMAGCL: A dual-masked adaptive graph contrastive learning framework for predicting circRNA-drug sensitivity. *Front. Genet.* 16:1721716. doi: 10.3389/fgene.2025.1721716

COPYRIGHT

© 2025 Wang, Guo, Li, Tang, Qi, Zhu and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

DMAGCL: A dual-masked adaptive graph contrastive learning framework for predicting circRNA-drug sensitivity

Peng Wang¹, Yuqi Guo¹, Zejun Li^{2*}, Di Tang¹, Mingming Qi^{3*}, Zhanyi Zhu⁴ and Lichao Zhang^{5*}

¹School of Electronic Information, Hunan First Normal University, Changsha, Hunan, China, ²School of Computer and Information Science, Hunan Institute of Technology, Hengyang, China, ³School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou, China, ⁴School of Software Quanzhou University of Information Engineering, Quanzhou, China, ⁵School of Intelligent Manufacturing and Equipment, Shenzhen University of Information Technology, Shenzhen, China

Circular RNAs (circRNAs) are a unique class of non-coding RNAs with stable covalently closed structures that play key regulatory roles in gene expression and drug response. However, experimental identification of circRNA-drug sensitivity remains labor-intensive. To overcome these limitations, we introduce DMAGCL, a Dual-Masked Graph Contrastive Learning framework, whose core innovations include: (1) a synergistic dual-masking strategy (path- and edge-level) that forces the model to learn robust representations against both macro-level path disruptions and micro-level edge noise; (2) an adaptive contrastive loss with a scheduled temperature parameter (t) to dynamically balance exploration and exploitation during training; and (3) an attention-based fusion classifier (AFC) that explicitly models complex cross-modal interactions between circRNA sequences and drug molecular graphs for adaptive multi-source information fusion. Comprehensive evaluations demonstrate that DMAGCL achieves state-of-the-art performance, attaining an average AUC of 0.8940 and AUPR of 0.9006 under five-fold cross-validation, and a slightly higher average AUC of 0.8982 under the more stringent ten-fold cross-validation, consistently surpassing strong baselines including GATECDA and MNGACDA. This performance advantage stems from our core design choices, as evidenced by systematic ablation studies confirming the indispensable and complementary roles of the dual-masking strategy and the effectiveness of the adaptive loss and fusion classifier. Case studies on four representative anticancer drugs (doxorubicin, gefitinib, sorafenib, and paclitaxel) achieved an average experimental validation rate of 80%, highlighting the framework's predictive reliability and biological relevance. In conclusion, this study makes three primary contributions: (1) it introduces the novel DMAGCL framework, establishing a new paradigm for circRNA-drug association prediction via its synergistic dual-masking, adaptive learning, and attentive fusion components; (2) it delivers a highly robust and interpretable model with validated predictive reliability through extensive experiments and case studies (80% average validation rate); and (3) it provides a scalable computational tool that offers valuable insights for

discovering novel circRNA-drug associations, understanding drug resistance mechanisms, and informing precision therapy design, with clear pathways for extension to other biological interaction tasks.

KEYWORDS

attention fusion classifier, circRNA, drug sensitivity prediction, dual-masked graph contrastive, multi-source feature integration

1 Introduction

Circular RNAs (circRNAs) exist stably through their covalently closed structure and have been proven to be key regulatory factors for tumor drug resistance (Chen and Yang, 2015; Suzuki and Tsukahara, 2014). As competitive endogenous RNAs, they can “sponge” miRNAs through miRNA response elements and thereby interfere with post-transcriptional silencing of mRNAs (Min et al., 2019). However, the complete molecular network of circRNA-drug sensitivity (CDS) remains to be elucidated. For instance, miRNAs can mediate post-transcriptional silencing by binding to the 3′-untranslated region of target mRNAs, while circRNAs can antagonize this process through molecular sponge effects (Bartel, 2004). circRNAs are pivotal regulators of tumor drug resistance (Min et al., 2019), yet computational elucidation of their interactions with drug sensitivity remains challenging (Bartel, 2004). Multiple studies have indicated that circHIPK3, Circ_0006528, etc. participate in tumor occurrence by regulating apoptosis and proliferation pathways (Hanahan and Weinberg, 2000), and can serve as diagnostic or prognostic markers (Li et al., 2018). At the level of drug resistance, circ_0076305 is upregulated by miR-186-5p to induce ABCC1 expression, thereby enabling non-small cell lung cancer to tolerate cisplatin (Liu et al., 2018); while circ_0072083 enhances the resistance of glioma to temozolomide through the miR-1252-5p/ALKBH5/NANOG axis (Wang et al., 2023). Although functional evidence is constantly emerging, experimental identification of CDS associations remains time-consuming and labor-intensive, and there is an urgent need for efficient computational frameworks to systematically analyze their network patterns. Additionally, in bladder cancer, CircNR3C1 can bind to BRD4 and interfere with the formation of the oncogenic complex between C-myc, thereby inhibiting tumor progression (Ding et al., 2021); studies have also shown that the ectopic expression of C-myc can partially reverse the tumor suppressive effect of this circRNA *in vivo* (Xie et al., 2020). These findings systematically reveal the important regulatory functions of circRNAs in tumor resistance. However, the complete molecular network of their interaction with drug sensitivity still needs to be further elucidated. Extensive experiments demonstrate that DMCL achieves state-of-the-art performance. Case studies on four anti-cancer drugs show an average experimental validation rate of 80%, underscoring its predictive reliability and biological relevance. This work provides a powerful computational tool and offers new insights into non-coding RNA mechanisms in drug resistance.

In the circRNA and drug sensitivity studies, the statistical models based on correlations play a crucial role. The core of these models lies in using computational methods to predict and uncover potential circRNA-drug sensitivity correlations, thereby effectively avoiding the time-consuming and labor-intensive drawbacks of traditional biological experiments. Computational

models offer an efficient approach for large-scale identification of such associations. Existing methods primarily rely on strategies such as graph neural networks (GNNs), multi-source feature integration, or random walks. Multi-source biological information integration and machine learning algorithms have become the mainstream approach for predicting drug sensitivity of circRNAs (Wei et al., 2023a). By introducing graph regularization into protein language models, it has been verified that graph constraints can improve the prediction of TCR-epitope binding specificity (Fu et al., 2025b); Zhou et al. proposed a global-local dual perspective to achieve multi-scale feature fusion of drug-protein interactions (Zhou et al., 2024b). The multi-source feature framework emphasizes the crucial regulatory role of circular RNAs in drug sensitivity and efficacy (Yin et al., 2024), while MNCLCDA reveals the association between circular RNAs and drug resistance and tumor progression through mixed neighborhood contrast learning (Li et al., 2023). Given the natural advantages of graph neural networks in complex biological networks, studies usually combine attention mechanisms to strengthen key features (Ru et al., 2022): MNGACDA uses graph autoencoders to construct multimodal networks (Yang and Chen, 2023), DeepWalk-GAT combines convolutional neural networks to extract topological sequence information (Li et al., 2023), Graph Attention Autoencoder (Deng et al., 2022), HETACDA (Xiao et al., 2023), Double-Layer Multi-Core Attention Network (Lu et al., 2023), and multimodal graph representation framework (Liu Z. et al., 2025) all effectively predict circular RNA-drug sensitivity through computational means, compensating for the lack of experimental throughput (Ru et al., 2021a). However, these methods have inherent limitations: GNN models are susceptible to information redundancy and noise within networks, and their decision-making process often lacks transparency. Similarity-based methods, on the other hand, struggle to capture complex non-linear biological relationships. To address these challenges, we propose a double-mask contrastive learning framework. The core innovation involves the simultaneous application of a path mask and an edge mask. The path mask is designed to deconstruct high-order semantic information in heterogeneous circRNA-drug relationships, while the edge mask enhances the model's robustness to local critical interactions. This dual design effectively suppresses information redundancy and noise in biological networks through contrastive learning. Furthermore, we introduce a dynamic temperature parameter strategy that adaptively optimizes the learning process based on the training state. Random walks and similarity measures are intertwined, forming a network that, through topological propagation, triggers potential correlations: an asymmetric dual-walk strategy re-evaluates the connection between circular RNAs and diseases, providing algorithmic support for rapid target identification (Toprak, 2024). The multi-head attention mechanism is used to convert the potential vocabulary between non-coding RNAs and

proteins into interaction codes (Zhou et al., 2023). Graph collaborative filtering and multi-view contrastive learning are combined, enabling miRNAs to confront their own sensitivity to drugs in the generated embedding space (Wei et al., 2023b). Pairwise learning guides dual graph convolution to shape a clearer interface representation of ncRNA-protein pairs (Zhuo et al., 2022), while the head-tail sampling protocol extracts key nodes from the sparse interaction wasteland, providing replenishment for downstream GNNs (Wei et al., 2023a). Benchmarking and framework parallel development: an evaluation standard for circular RNA-disease prediction was established and quietly expanded to the field of drug sensitivity (Lan et al., 2023); through the method of extracting graph skeletons and fusing attention, researchers gradually revealed the progressive process of non-coding RNA-mediated drug resistance (Zhang et al., 2023); a comprehensive study revealed how deep learners locate binding sites of RNA-binding proteins on circular RNAs and how this binding implicitly regulates drug responses (Wang Z. et al., 2025). Multidimensional data convergence, complex algorithms take off, the statistical-graph hybrid now charts a high-throughput, low-cost route for drug sensitivity based on circular RNAs, accelerating the transformation of potential biomarkers and therapeutic targets from the realm of stars to clinical reality (Ru et al., 2021b).

Network models have emerged as a pivotal tool in bioinformatics and systems biology for modeling complex biological processes and diseases. For instance, Peng et al. developed an explainable multi-scale framework for circRNA-miRNA interaction prediction, highlighting the utility of multi-scale feature engineering in circRNA research (Peng et al., 2025b). In another study, Peng et al. introduced metaCDA, a meta-learning framework for circRNA-driven drug discovery, showcasing the potential of adaptive aggregation and meta-knowledge in circRNA studies (Peng et al., 2025a). By modeling biological entities and their interactions as graphs, network science has been widely applied in protein-protein interaction networks, gene regulatory networks, and brain connectomes. For instance, Milan and Canataro applied network inference to the research of cancer and neurodegenerative diseases (Milano and Cannataro, 2023); through the heterogeneous GCN combined with pseudo-path bidirectional attention (Niu et al., 2025), NSL2CD's adaptive subspace embedding (Xiao et al., 2021b), AAECDA's MSCNN-adversarial autoencoder pipeline (Wang Y. et al., 2025), and Xiao's propagation, path, matrix and depth-based classification framework (Xiao et al., 2021a), as well as the cross-modal ternary attention in ET-PROTACs (Cai et al., 2025), Kolmogorov-Arnold drug KANs (Fu et al., 2025a), and masked path GAM-MDR (Zhou et al., 2024a), they jointly advanced the frontier of topological research. These studies have jointly promoted the development of circRNA-drug sensitivity association prediction based on network topological structures, providing new perspectives and methods for biomarker identification and drug development in complex diseases.

Machine learning is playing an increasingly important role in drug discovery and prediction, especially in drug sensitivity prediction. In the field of predicting drug sensitivity for cancer cell lines, even when the training samples are limited, traditional molecular fingerprints have been surpassed or even equalled by end-to-end TextCNN (Baptista et al., 2022). Breakthroughs have also been frequently observed in the circRNA domain, such as

LSNSCDA, which overcomes the limitations of fixed step size and negative sample noise through a local smoothing graph neural network and reliable negative sampling (Fan et al., 2024); MAGSDMF framework, which precisely captures the potential circRNA-drug sensitivity associations through multiple attention, graph learning, and deep matrix decomposition (Ai et al., 2024); DCDA uses a feedforward-self-encoding hybrid to integrate multi-source data to shape circRNA-disease features and performs exceptionally well (Turgut et al., 2023); SNFTPGd-CDA achieves high AUC in nonlinear fusion of multi-source information through similar network fusion-tensor product graph diffusion parallel cascaded forest (Liu et al., 2024); JLCRB multi-view collaborative representation network enhances cross-view consistency to locate circRNA binding sites, and the average AUC has reached a new high (Du and Xue, 2022). Looking at the entire process of drug development, ModDRDSP uses deep bidirectional GRU and message passing networks to jointly depict multimodal drug information, and then integrates cell line multi-omics data, and is completed by a deep forest for sensitivity prediction, with the model's performance comprehensively leading existing methods (Song et al., 2024).

In the research on the association between circRNAs and drug sensitivity prediction, statistical models based on correlations, models based on graph neural networks (GNN) and attention mechanisms, as well as models based on random walks and similarity measures each have their own advantages and limitations. The statistical model based on 155-dimensional correlation coefficients integrates multi-source biological data and employs machine learning algorithms to construct a prediction framework. It can simultaneously capture correlation signals in multiple dimensions such as expression profiles, drug structures, and clinical phenotypes, thereby enhancing prediction coverage and accuracy (Chen et al., 2022). Due to its high dependence on data quality and completeness, this method is prone to overfitting when the training samples are insufficient or the feature dimensions are excessive, resulting in a decline in generalization performance. In contrast, the combination of graph neural networks and attention mechanisms utilizes topological structure information for embedding learning of circRNAs and drugs, highlighting key node features through dynamic weights, demonstrating strong association mining capabilities in large-scale network scenarios (Li X. et al., 2025). However, this strategy has high computational costs, high memory usage, and a large number of hyperparameters, resulting in significant tuning costs. With the random walk and similarity measurement methods constructing a similarity network of biological entities and performing information propagation on it, potential associations can be rapidly scanned with linear complexity, suitable for initial screening tasks of large datasets (Li M. et al., 2025). Its performance is highly dependent on the accuracy of similarity measurement, and because the model is essentially a linear method, it is difficult to depict the nonlinear interactions within the biological system, resulting in insufficient performance in complex association prediction scenarios. Moreover, during the process of dimensionality reduction or network embedding, important biological details may be lost, affecting the precision of the prediction.

In summary, these three types of methods have distinct characteristics and complement each other in the prediction of the association between circRNA and drug sensitivity. Future

research can integrate the advantages of these methods, such as combining the interpretability enhancement technology of GNN and the integration of multi-source features, to improve the prediction accuracy and biological relevance (Ma et al., 2024). At the same time, introducing data augmentation techniques, interpretability tools, and establishing standardized benchmark test datasets and evaluation indicators will help further optimize the model performance and promote the progress of this field (Liu T. et al., 2025). Moreover, exploring the application of these methods in other biomedical fields, such as gene regulatory networks and protein interaction networks, will further verify their universality and effectiveness.

2 Methods

We put forward a novel framework entitled DMAGCL. This framework is specifically designed to concurrently tackle three crucial challenges during the process of graph representation learning: information redundancy, negative sample noise, and class imbalance issues. Through an innovative architectural design, our method integrates graph encoding, masked reconstruction, dynamic negative sampling, ensemble learning, and ranking optimization into a unified end-to-end learning objective. This integration allows for the coordinated evolution of pseudo-label generation, sampling distribution adjustment, model ensemble, and evaluation metric optimization throughout the training process. The overall process is shown in Figure 1.

2.1 Multi-source feature extraction

To comprehensively characterize the characteristics of nodes in the network, we systematically extracted multi-source features from two dimensions: circRNAs and drugs. We then constructed a unified feature representation through an effective fusion strategy, providing high-quality feature inputs for subsequent graph representation learning. For circRNA nodes, we integrated two complementary feature sources. Based on the nucleic acid sequence information of circular RNA host genes, we used the edit distance algorithm to calculate the similarity between sequences, and constructed a circRNA sequence similarity matrix $\mathbf{R}_{\text{seq}} \in \mathbb{R}^{M \times M}$. The elements of this matrix are defined as:

$$\mathbf{R}_{\text{seq}}(i, j) = 1 - \frac{\mathcal{D}_{\text{edit}}(\text{Seq}_i, \text{Seq}_j)}{\max(|\text{Seq}_i|, |\text{Seq}_j|)}, \quad (1)$$

Here, $\mathcal{D}_{\text{edit}}$ represents the minimum number of editing operations required to transform one sequence into another, and Seq_i denotes the host gene sequence of the i -th circRNA. Additionally, based on the known circRNA-drug sensitivity correlation matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, we constructed the Gaussian interaction spectral kernel similarity matrix $\mathbf{R}_{\text{gip}} \in \mathbb{R}^{M \times M}$.

$$\mathbf{R}_{\text{gip}}(i, j) = \exp(-\eta_c \|\mathbf{A}(i, :) - \mathbf{A}(j, :)\|^2), \quad (2)$$

The nuclear bandwidth parameter η_c is determined by the following formula.

$$\eta_c = \left(\frac{1}{M} \sum_{k=1}^M \|\mathbf{A}(k, :)\|^2 \right)^{-1}. \quad (3)$$

Ultimately, we obtained the comprehensive feature matrix $\mathbf{X}_c \in \mathbb{R}^{M \times M}$ of the circular RNA through conditional weighted fusion: when sequence similarity is available, the average value of the two similarities is taken; otherwise, only the Gaussian interaction spectral similarity is retained. In terms of the construction of drug characteristics, we adopted a similar dual-source feature fusion strategy. Based on the chemical structure information of the drugs, we used topological fingerprints and Tanimoto coefficients to calculate the structural similarity between drugs, forming a drug structure similarity matrix $\mathbf{D}_{\text{stru}} \in \mathbb{R}^{N \times N}$.

$$\mathbf{D}_{\text{stru}}(i, j) = \frac{|F(d_i) \cap F(d_j)|}{|F(d_i) \cup F(d_j)|}, \quad (4)$$

Here, $\mathcal{F}(d_i)$ represents the topological fingerprint feature vector of drug d_i . Meanwhile, based on the same circRNA-drug sensitivity association matrix $\mathbf{A}$, we constructed the Gaussian interaction spectral kernel similarity matrix $\mathbf{D}_{\text{gip}} \in \mathbb{R}^{N \times N}$,

$$\mathbf{D}_{\text{gip}}(i, j) = \exp(-\eta_d \|\mathbf{A}(:, i) - \mathbf{A}(:, j)\|^2) \quad (5)$$

The parameter η_d is determined by the following formula:

$$\eta_d = \frac{1}{N} \sum_{k=1}^N \|\mathbf{A}(:, k)\|^2 \quad (6)$$

The similarity features of the two drugs are fused through the same weighted strategy under the same conditions, resulting in a comprehensive feature matrix $\mathbf{X}_d \in \mathbb{R}^{N \times N}$ for the drugs. Based on the above feature matrix and the known correlation between circRNAs and drug sensitivity, we constructed a heterogeneous network $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V} = \mathcal{C} \cup \mathcal{D}$ includes all circRNA and drug nodes, and the edge set $\mathcal{E}$ integrates three types of connection relationships: circRNA - circRNA similarity edges, drug - drug similarity edges, and circRNA - drug association edges.

2.2 Dual masked graph views

While graph structure is vital for representation learning, standard GNNs can be limited in capturing complex topological dependencies. To address this, we propose a self-supervised contrastive learning framework with a dual-masking strategy to enhance the robustness and quality of graph representations. This method constructs two masking views with different structural perturbation intensities, capturing local structural patterns at the path level and edge level respectively, and leveraging the contrastive learning mechanism to improve the model's perception of graph structure information.

Specifically, we designed two complementary masking modules: the Path Masking Module (MaskPath) and the Edge Masking Module (MaskEdge). The Path Masking Module randomly samples the node paths in the graph and performs overall masking on the nodes and their associated edges within the path, thereby generating the first perturbed view $\mathcal{G}_1$. This operation can

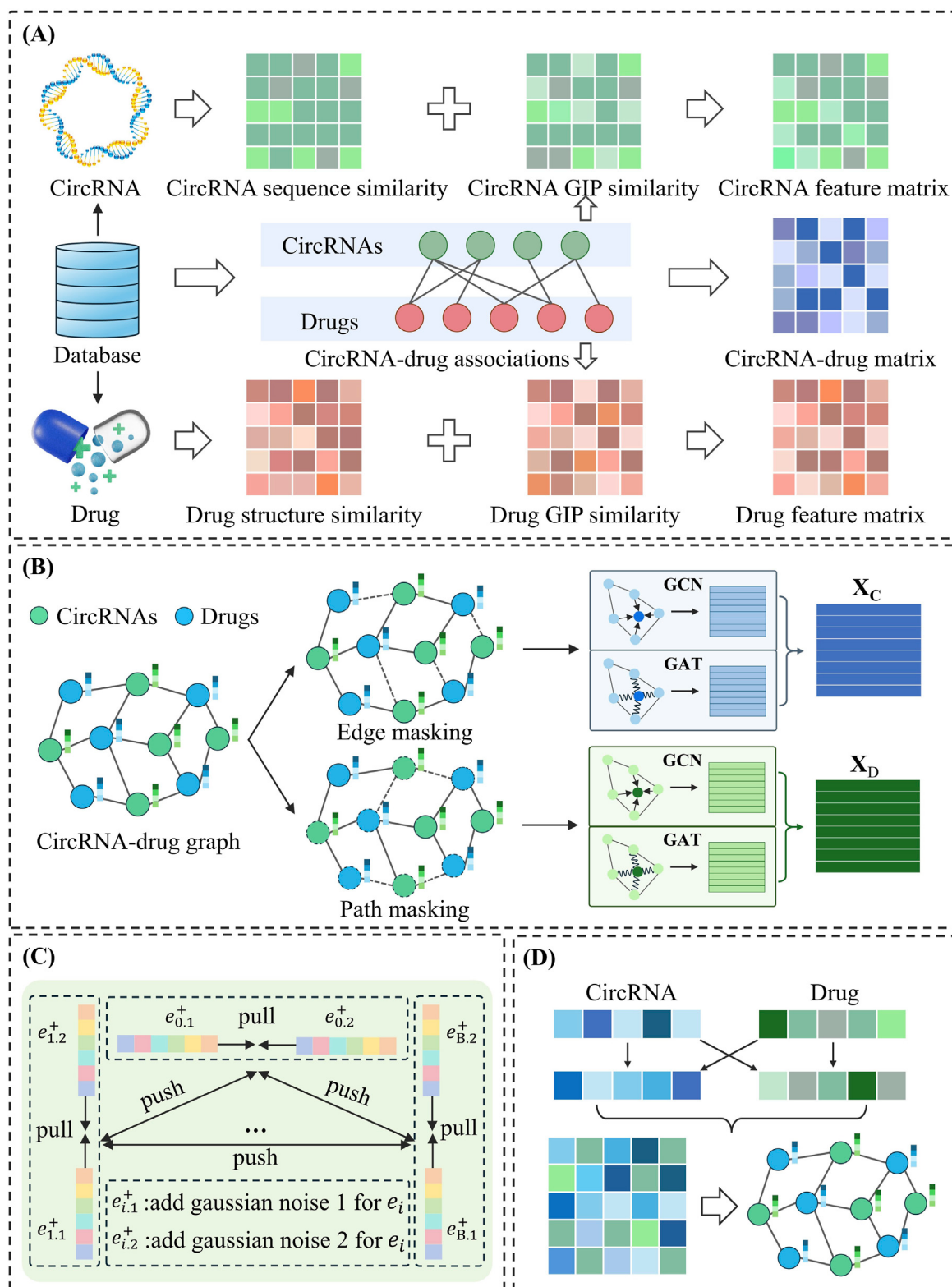


FIGURE 1
The computational flowchart of DMAGCL. (A) CircRNA and drug feature extraction (B) Dual view masking graph learning. (C) Adaptive contrastive learning (D) CircRNA-drug graph reconstruction.

simulate the absence of local connectivity patterns in the graph, forcing the model to still infer the dependencies between nodes even when the paths are incomplete. The Edge Masking Module is more

granular, randomly selecting some edges in the graph for masking, generating the second perturbed view $\mathcal{G}_2$. This operation aims to weaken the local direct connections, enabling the model to pay more

attention to other structural cues within the neighborhood. We introduce two complementary masking modules: MaskPath and MaskEdge. MaskPath samples and masks entire node paths, creating view $\mathcal{G}_1$ and compelling the model to recover missing connectivity patterns. MaskEdge operates at a finer granularity by randomly dropping individual edges to form view $\mathcal{G}_2$, forcing the model to rely on broader neighborhood cues. This dual approach ensures that the model learns robust representations against both macro-level path disruptions and micro-level edge noise. The two masking operations can be formally represented as:

$$\mathcal{G}_1 = \text{MaskPath}(\mathcal{G}, \alpha) \quad (7)$$

$$\mathcal{G}_2 = \text{MaskEdge}(\mathcal{G}, \beta) \quad (8)$$

Among them, α and β are the intensity control parameters for the path mask and edge mask, respectively. They are used to adjust the range and degree of the masks, thereby controlling the disturbance intensity and diversity of the view.

We employ a shared-weight graph encoder $\mathcal{E}$ to map both masked views to a unified representation space. The encoder processes each view independently, generating corresponding graph-level representations $\mathbf{Z}_1$ and $\mathbf{Z}_2$:

$$\mathbf{Z}_1 = \mathcal{E}(\mathcal{G}_1) \quad (9)$$

$$\mathbf{Z}_2 = \mathcal{E}(\mathcal{G}_2) \quad (10)$$

The encoder $\mathcal{E}$ can be implemented by combining a graph convolutional network (GCN) and a graph attention network (GAT). After obtaining the view representations, we use the adaptive contrastive loss as the contrastive target. By maximizing the mutual information between the representations of the same graph under different masking strategies, the model is guided to learn representations that are robust to structural perturbations. During the model training process, the input graph $\mathcal{G}$ is first subjected to path masking and edge masking to generate two complementary perturbed views $\mathcal{G}_1$ and $\mathcal{G}_2$. Subsequently, the shared graph encoder extracts the graph-level representations $\mathbf{Z}_1$ and $\mathbf{Z}_2$ of the two. Next, the similarity of the representations is calculated, and a contrastive learning task is constructed based on the loss. Finally, the encoder parameters are optimized through the backpropagation mechanism to minimize the contrastive loss. This training strategy enables the model to extract common structural information from different masking perspectives, thereby effectively enhancing the modeling ability of local neighborhood dependencies and improving the quality and generalization of graph structure representations.

2.3 Multi-view graph encoder

In the field of graph neural networks, the effective modeling and feature extraction of multi-view graph data has always been a key and challenging research direction. Since multi-view graph data typically contains heterogeneous structural information from different sources or different feature spaces, how to fully integrate the effective features from each view and retain their inherent topological relationships is an important issue for improving the representational learning ability of graphs. To effectively integrate heterogeneous information from our constructed multi-view graph (e.g., sequence similarity, GIP

similarity), we propose a Multi-View Graph Encoder. This encoder combines GCN for intra-view feature extraction and GAT for inter-view fusion, ensuring a balanced and comprehensive representation. The core design of this method consists of two main stages: local structure modeling within views and global feature fusion between views. Firstly, for the structural characteristics within each view, GCN is used to aggregate the neighborhood of nodes to capture the specific local topological patterns of the view. Specifically, for the v -th view, the node feature matrix is $\mathbf{X}^{(v)}$ and the adjacency matrix is $\mathbf{A}^{(v)}$. To enhance training stability and retain the node's own features, self-loops are introduced to construct an augmented adjacency matrix $\tilde{\mathbf{A}}^{(v)} = \mathbf{A}^{(v)} + \mathbf{I}$, and the corresponding degree matrix $\tilde{\mathbf{D}}$ is calculated. The node feature update formula in the GCN layer is:

$$\mathbf{H}^{(v)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}}^{(v)} \tilde{\mathbf{D}}^{-1/2} \mathbf{X}^{(v)} \mathbf{W}^{(v)}\right), \quad (11)$$

Here, $\mathbf{W}^{(v)}$ represents the trainable weight matrix specific to the view, and σ is the nonlinear activation function. Through this step, the smooth propagation of node features within each view and the structured-aware representation learning can be achieved.

After obtaining the node representations of each view, in order to further integrate the information from multiple views, this paper introduces the GAT mechanism to adaptively learn the contribution weights of different views to the current view. Specifically, for the target view v , its output representation after integrating the features of other views is:

$$\mathbf{H}_{\text{attn}}^{(v)} = \sigma\left(\sum_{u=1}^V \alpha_{vu} \mathbf{H}^{(u)}\right) \quad (12)$$

Here, α_{vu} represents the attention weight from view u to view v , reflecting the significance of view u in supplementing the features of view v . The calculation of this weight employs the following attention mechanism:

$$\alpha_{vu} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{H}^{(v)} \parallel \mathbf{H}^{(u)}]))}{\sum_{u=1}^V \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{H}^{(v)} \parallel \mathbf{H}^{(u)}]))} \quad (13)$$

Here, $\mathbf{a}$ represents a learnable attention parameter vector, and the symbol $\parallel$ denotes the feature concatenation operation. Through this mechanism, the model can automatically identify the complementarity among different views and endow the information fusion process with greater interpretability.

Finally, the feature representations of each view, weighted by attention, are integrated and input into a fully connected layer or a specific downstream task module to generate the final graph representation or node representation. The integrated output can be expressed as:

$$\mathbf{Z} = \sigma(\mathbf{W} [\mathbf{H}_{\text{attn}}^{(1)}, \mathbf{H}_{\text{attn}}^{(2)}, \dots, \mathbf{H}_{\text{attn}}^{(V)}]) \quad (14)$$

Here, $\mathbf{W}$ represents the globally trainable weight matrix, and σ is the activation function of the output layer. This integration strategy can integrate the advantages of each view, enhancing the model's ability to model complex graph structures and multi-source information.

In conclusion, the multi-view graph encoder proposed in this paper combines the advantages of GCN in modeling the internal

structure of views with the flexibility of GAT in integrating views, which not only effectively extracts the local features of each view, but also captures the global correlations between views. Thus, it achieves more comprehensive and robust feature learning in multi-view graph data processing.

2.4 Adaptive contrastive loss

As a robust and effective alternative, we introduce a Scheduled Contrastive Loss featuring a dynamic temperature parameter (t) that follows a predefined linear decay schedule. Specifically, this mechanism sets a higher temperature value in the early training stage to encourage the model to conduct more extensive exploration and avoid prematurely settling into suboptimal solutions; while gradually reducing the temperature value in the later training stage to enhance the model's ability to distinguish between positive and negative samples, thereby improving its generalization performance.

The mathematical definition of the adaptive contrastive loss function is as follows. Let the input feature vectors be $\mathbf{z}_i$ and $\mathbf{z}_j$, where i and j represent the positive sample pairs, and the other samples k (where $k \neq i, j$) are regarded as negative samples. The loss function can be expressed as:

$$\mathcal{L}_{\text{adaptive}} = -\log \frac{\exp(\frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\tau(t)})}{\sum_{k=1}^K \exp(\frac{\mathbf{z}_i \cdot \mathbf{z}_k}{\tau(t)})} \quad (15)$$

Here, $\mathbf{z}_i \cdot \mathbf{z}_j$ represents the dot product similarity between the feature vectors, K is the total number of negative samples, and $\tau(t)$ is a dynamic temperature parameter that adjusts according to the change in the training iteration number t . The definition of the dynamic temperature parameter $\tau(t)$ is:

$$\tau(t) = \tau_{\max} - \frac{t}{T} (\tau_{\max} - \tau_{\min}) \quad (16)$$

Here, $\tau_{\max}$ (initial temperature) encourages broad exploration early in training by smoothing the loss landscape. $\tau_{\min}$ (final temperature) sharpens the focus on hard negatives later for better discrimination. T is the total number of iterations, governing the cooling schedule. This linearly decreasing temperature strategy simulates an 'annealing' process. It provides a structured learning curriculum for the model: starting with robust, noise-tolerant learning and progressively transitioning to high-precision discrimination. In the initial stage of training, a higher temperature enables the smooth handling of the similarity difference between positive and negative samples, thereby reducing the model's sensitivity to noise and encouraging more diverse feature learning; as training progresses, the temperature gradually decreases, and the model gradually focuses on precisely distinguishing positive and negative samples, which helps to improve the discriminative power of features.

During the training process, we first initialize the key parameters, including $\tau_{\max}$, $\tau_{\min}$ and the total number of iterations T . These parameters can be adjusted according to the data distribution of the specific task and the complexity of the model, for example, through cross-validation to optimize the selection. In each training iteration t , the temperature parameter

$\tau(t)$ is updated in real time using the formula $\tau(t) = \tau_{\max} - \frac{t}{T} (\tau_{\max} - \tau_{\min})$. Then, for each positive sample pair $(\mathbf{z}_i, \mathbf{z}_j)$, the similarity matrix between it and all other samples (including positive and negative samples) is calculated, and the contrastive loss $\mathcal{L}_{\text{adaptive}}$ is computed using the current dynamic temperature $\tau(t)$. The model parameters are updated using the backpropagation algorithm to minimize the loss function, while monitoring the training loss and validation set performance to avoid overfitting. This dynamic temperature adjustment mechanism not only effectively alleviates the problems of premature convergence and overfitting caused by fixed temperature parameters, but also adaptively balances the exploration and exploitation phases of the model, thereby improving the training efficiency while enhancing the generalization ability of the model on unseen data. Moreover, the flexibility of this method makes it easy to integrate into various contrastive learning frameworks, providing a scalable foundation for subsequent research.

2.5 Attention fusion classifier

Simple feature fusion (e.g., concatenation) fails to capture complex cross-modal interactions between circRNA sequences and drug molecular graphs. We therefore propose an Attention-based Fusion Classifier (AFC) to dynamically model these correlations and adaptively fuse multi-source information. Compared with traditional feature concatenation methods, AFC incorporates a multi-head attention structure, which can extract and fuse the interaction information between the circRNA sequence mode and the drug molecule graph mode from multiple subspaces, effectively enhancing the discriminative ability of the feature representation. Additionally, to further improve the robustness and generalization ability of the model, the classification layer adopts a hierarchical multi-layer perceptron (MLP) structure and combines batch normalization (BatchNorm) for regularization processing. Experiments have shown that AFC exhibits superior performance in multiple multimodal classification tasks.

Specifically, the core component of AFC is the multi-head attention mechanism. This mechanism can dynamically adjust the fusion weights based on the correlation between the input features of the circRNA sequence mode and the drug molecule graph mode, thereby focusing on the cross-modal information that is more helpful for the current task. Assuming the input is the feature embeddings of the circRNA sequence mode and the drug molecule graph mode, which are $\mathbf{F}_{\text{circ}} \in \mathbb{R}^{n \times d_{\text{circ}}}$ and $\mathbf{F}_{\text{drug}} \in \mathbb{R}^{n \times d_{\text{drug}}}$, where n is the number of samples and d_{circ} and d_{drug} are the feature dimensions of the corresponding modes. First, the features of the circRNA sequence mode and the drug molecule graph mode are linearly transformed to a unified latent space dimension d :

$$\mathbf{U} = \mathbf{F}_{\text{circ}} \mathbf{P}_U \in \mathbb{R}^{n \times d}, \quad \mathbf{S} = \mathbf{F}_{\text{drug}} \mathbf{P}_S \in \mathbb{R}^{n \times d}, \quad \mathbf{T} = \mathbf{F}_{\text{drug}} \mathbf{P}_T \in \mathbb{R}^{n \times d} \quad (17)$$

Among them, $\mathbf{P}_U \in \mathbb{R}^{d_{\text{circ}} \times d}$, $\mathbf{P}_S \in \mathbb{R}^{d_{\text{drug}} \times d}$, and $\mathbf{P}_T \in \mathbb{R}^{d_{\text{drug}} \times d}$ are learnable parameter matrices. Next, the query matrix $\mathbf{U}$, key matrix $\mathbf{S}$, and value matrix $\mathbf{T}$ are uniformly divided into h heads according to the feature dimension, with each head having a dimension of $d_h = \frac{d}{h}$. For each attention head i , the attention weight matrix $\mathbf{G}^i$ and

the corresponding fused feature $\mathbf{O}^i$ of this head are calculated separately:

$$\mathbf{G}^i = \text{softmax}\left(\frac{\mathbf{U}^i (\mathbf{S}^i)^T}{\sqrt{d_h}}\right) \in \mathbb{R}^{n \times n}, \quad \mathbf{O}^i = \mathbf{G}^i \mathbf{T}^i \in \mathbb{R}^{n \times d_h} \quad (18)$$

The attention weight $\mathbf{G}^i$ reflects the strength of the correlation between the circRNA sequence mode and the drug molecule graph mode across different samples, while $\mathbf{O}^i$ is the weighted combination of the drug molecule graph mode features guided by correlation. Subsequently, the outputs of all attention heads are concatenated along the feature dimension, and integrated through a linear mapping matrix $\mathbf{P}_O \in \mathbb{R}^{d \times d}$ to obtain the final fused representation:

$$\mathbf{C} = \text{Concat}(\mathbf{O}^0, \mathbf{O}^1, \dots, \mathbf{O}^{h-1}) \mathbf{P}_O \in \mathbb{R}^{n \times d} \quad (19)$$

Through the multi-head attention mechanism, AFC can jointly capture the complex dependencies between the circRNA sequence mode and the drug molecule graph mode from multiple representation subspaces, thereby more comprehensively integrating cross-modal information. In the classification module, AFC employs a hierarchical MLP structure to conduct deep semantic encoding of the fused features, and combines BatchNorm technology to enhance the training stability and generalization ability of the model. Specifically, the fused features $\mathbf{C}$ are first input into a two-layer MLP:

$$\mathbf{L}_1 = \text{ReLU}(\mathbf{C}\mathbf{A}_1 + \mathbf{c}_1) \in \mathbb{R}^{n \times d'}, \quad \mathbf{L}_2 = \text{ReLU}(\mathbf{L}_1\mathbf{A}_2 + \mathbf{c}_2) \in \mathbb{R}^{n \times d''} \quad (20)$$

Among them, $\mathbf{A}_1 \in \mathbb{R}^{d \times d'}$ and $\mathbf{A}_2 \in \mathbb{R}^{d' \times d''}$ are weight matrices, and $\mathbf{c}_1 \in \mathbb{R}^{d'}$ and $\mathbf{c}_2 \in \mathbb{R}^{d''}$ are bias terms. To alleviate the problem of gradient vanishing or internal covariate shift in deep networks, we apply batch normalization after the activation output of each layer of the MLP:

$$\mathbf{L}'_1 = \text{BatchNorm}(\mathbf{L}_1), \quad \mathbf{L}'_2 = \text{BatchNorm}(\mathbf{L}_2) \quad (21)$$

Ultimately, the normalized high-level features $\mathbf{L}'_2$ pass through a fully connected layer followed by a softmax activation function, resulting in the probability distributions corresponding to each class:

$$\mathbf{Y} = \text{softmax}(\mathbf{L}'_2\mathbf{A}_3 + \mathbf{c}_3) \in \mathbb{R}^{n \times C} \quad (22)$$

Among them, $\mathbf{A}_3 \in \mathbb{R}^{d'' \times C}$ and $\mathbf{c}_3 \in \mathbb{R}^C$ are the parameters of the classification layer, and C represents the number of categories. AFC combines the multi-head attention mechanism with the hierarchical regularization MLP structure, which not only enables dynamic and

adaptive fusion of circRNA sequence modalities and drug molecule graph modalities, but also maintains good expression ability and generalization performance in complex tasks.

3 Experiment

3.1 Experimental setup and parameter settings

To ensure the reproducibility of our experimental results, we provide a comprehensive description of our implementation details and hyperparameter settings. All experiments were conducted under the same computational environment, and the following key parameters were utilized throughout the study: Network Architecture: The number of layers for both the GCN and GAT in our multi-view graph encoder was set to 3. The hidden dimension of this encoder was configured to 32. Training Strategy: The model was trained for 300 epochs. The Adam optimizer was employed with a learning rate of 0.001. Regularization: A Dropout rate of 0.3 was applied to mitigate overfitting. Reproducibility: A fixed random seed of 42 was used to guarantee the stability of all experimental outcomes and to facilitate replication. The rationale behind the selection of these hyperparameters, including the learning rate and model depth, is further supported by the sensitivity and ablation studies presented in subsequent sections.

3.2 Performance comparison

The five-fold cross-validation results, detailed in Table 1, demonstrate the remarkable stability of our model across different data partitions. The AUC values across all five folds exhibited minimal fluctuation, ranging narrowly from 0.8918 to 0.8962, yielding a high average AUC of 0.8940. This consistency, mirrored in other metrics such as AUPR, F1-score, Accuracy, and Recall, indicates that the model's performance is not dependent on a particular split of the data and reliably reproduces the performance of the complete model. This low variance is a strong indicator of the model's robustness. Other indicators, including AUPR, F1-score, Accuracy, and Recall, also showed similar consistency. This indicates that the model has good adaptability to different training-test data partitions. To further stress-test the model's robustness, we conducted a more stringent ten-fold cross-validation. The results, presented in Table 2, reveal an even

TABLE 1 The results of 5-fold cross-validation.

Metrics/Fold	AUC	AUPR	F1-score	Accuracy	Recall
1	0.8962	0.9028	0.8371	0.8332	0.8556
2	0.8923	0.8991	0.8330	0.8287	0.8502
3	0.8951	0.9015	0.8358	0.8319	0.8537
4	0.8918	0.8983	0.8319	0.8276	0.8489
5	0.8946	0.9013	0.8357	0.8311	0.8527
Average	0.8940	0.9006	0.8347	0.8305	0.8523

TABLE 2 The results of 10-fold cross-validation.

Fold	AUC	AUPR	F1-score	Accuracy	Recall
1	0.9001	0.8994	0.8385	0.8359	0.8551
2	0.8968	0.8953	0.8332	0.8302	0.8493
3	0.8974	0.8961	0.8341	0.8313	0.8504
4	0.8992	0.8986	0.8376	0.8352	0.8542
5	0.8962	0.8949	0.8324	0.8294	0.8486
6	0.8986	0.8979	0.8363	0.8339	0.8527
7	0.8978	0.8966	0.8350	0.8321	0.8515
8	0.8990	0.8982	0.8369	0.8346	0.8534
9	0.8965	0.8943	0.8321	0.8290	0.8481
10	0.8984	0.8970	0.8363	0.8333	0.8524
Average	0.8982	0.8973	0.8354	0.8329	0.8517

higher level of consistency across ten different data splits. The average AUC improved slightly to 0.8982, with values ranging from 0.8962 to 0.9001. This minor performance improvement under a more granular data division suggests that the model can effectively leverage larger training subsets. The consistently high performance and low standard deviations across all key metrics (AUPR, F1-score, etc.) in this setting provide compelling statistical evidence for the model's generalizability and stability. The AUC values ranged from 0.8962 to 0.9001, with an average of 0.8982, and this result was even slightly higher than that of the five-fold cross-

validation. Other key indicators such as AUPR (average 0.8973), F1-score (average 0.8354), Accuracy (average 0.8329), and Recall (average 0.8517) also remained at a high level, and the standard deviations between each fold were all within a relatively small range.

Figure 2 can more intuitively show the comparison of the results of DMAGCL with the baseline model 5-fold and 10-fold cross-validation. The results of cross-validation fully demonstrate from a statistical perspective that the model has excellent generalization ability and stability. Under different data divisions, the model can maintain excellent performance, indicating that our method not only has a good fitting ability for the training data, but also has reliable predictive performance for unknown data.

We evaluated our dual-mask graph contrastive learning framework against several representative methods for circRNA-drug sensitivity association prediction. The baselines included traditional machine learning models (SVM, RF, KNN, AdaBoost) and recent graph neural networks (VGAE, VGAMF, GCNMDA, GATECDA, MNGACDA). All experiments were conducted under the same experimental environment and dataset, using five-fold and ten-fold cross-validation to ensure the statistical reliability of the results.

A comprehensive performance comparison against various baseline methods under five-fold cross-validation is summarized in Table 3. Our proposed method consistently outperforms all competitors, achieving the top scores across all five evaluation metrics. Notably, it attains an AUC of 0.8940 and an AUPR of 0.9006, representing a clear margin over the strongest baselines. The results reveal a key trend: graph neural network-based methods

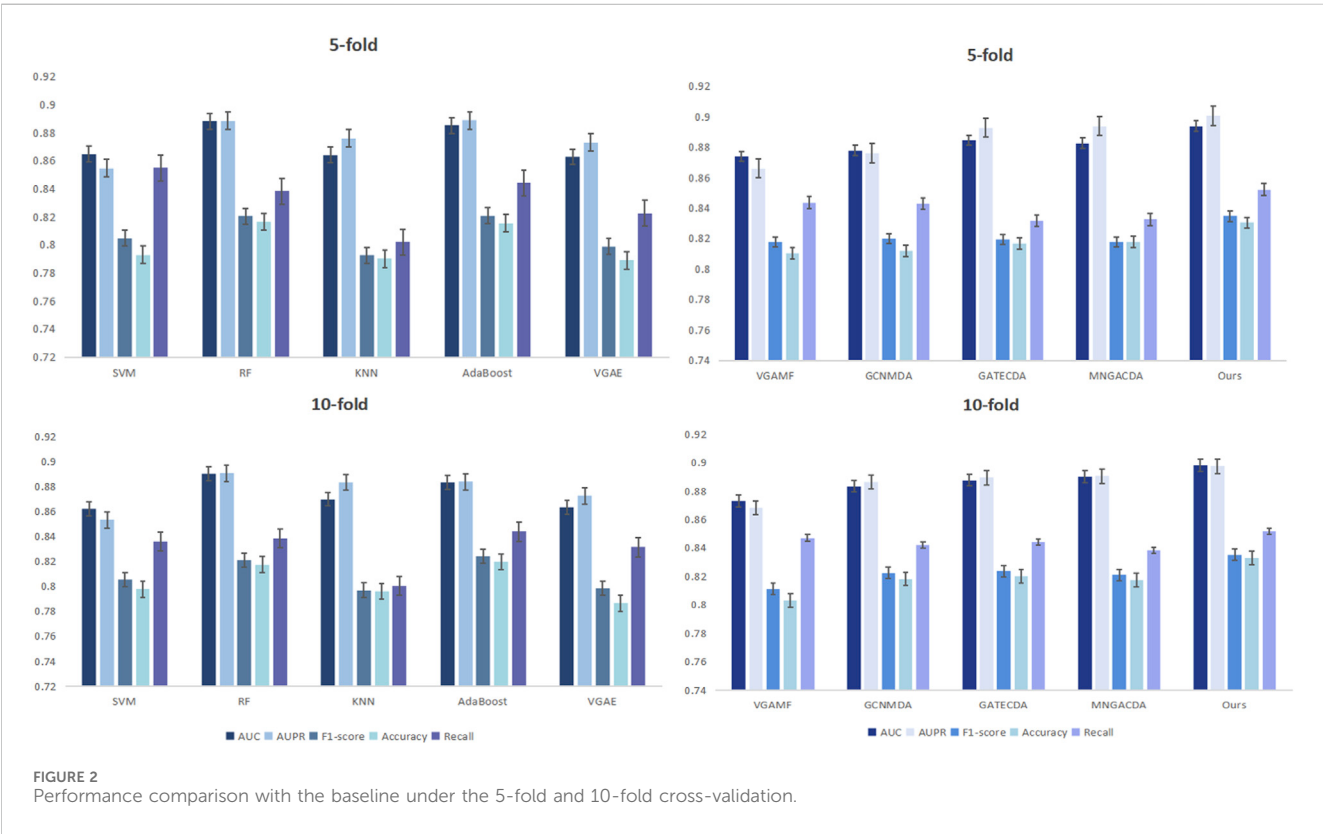


TABLE 3 Performance comparison with the baseline under the 5-fold cross-validation.

Methods	AUC	AUPR	F1-score	Accuracy	Recall
SVM	0.8648	0.8547	0.8049	0.7928	0.8550
RF	0.8881	0.8885	0.8204	0.8165	0.8383
KNN	0.8642	0.8760	0.7926	0.7901	0.8020
AdaBoost	0.8852	0.8888	0.8207	0.8155	0.8443
VGAE	0.8628	0.8730	0.7988	0.7892	0.8227
VGAMF	0.8740	0.8662	0.8176	0.8104	0.8437
GCNMDA	0.8778	0.8762	0.8198	0.8119	0.8428
GATECDA	0.8846	0.8929	0.8194	0.8168	0.8316
MNGACDA	0.8826	0.8940	0.8178	0.8178	0.8326
Ours	0.8940	0.9006	0.8347	0.8305	0.8523

(e.g., GATECDA, MNGACDA) generally surpass traditional machine learning models (e.g., RF, SVM), underscoring the critical importance of explicitly modeling the graph structure for this prediction task. However, even among these advanced graph methods, our dual-mask framework establishes a new state-of-the-art. Including AUC, AUPR, F1-score, accuracy, and recall rate. The AUC reached 0.8940, and AUPR reached 0.9006, significantly outperforming other comparison methods. Traditional methods like RF and AdaBoost showed competitive results but were limited in capturing complex graph topology. Graph-based methods generally outperformed them, underscoring the importance of graph structure modeling. Notably, even attention-based methods like GATECDA and MNGACDA were outperformed by our framework, suggesting that our dual-mask strategy and contrastive learning paradigm better capture the intricate circRNA-drug associations than attention mechanisms alone.

The superiority of our method is further confirmed under the more granular ten-fold cross-validation, as detailed in Table 4. Here, our model maintains its leading advantage with an average AUC of 0.8982 and AUPR of 0.8973. It is worth noting that while some baselines like GATECDA also showed improved performance with more folds, the performance gap between our method and these strong competitors persists. This consistent outperformance across both five-fold and ten-fold validations strongly suggests that the gains from our dual-mask contrastive learning framework are robust and not merely an artifact of a specific evaluation setup. Other indicators also show a similar trend. Compared with the five-fold cross-validation, the performance fluctuations of each method under the ten-fold validation are smaller, indicating that the method proposed in this paper has better adaptability to different data divisions. It is worth noting that in the ten-fold validation, the performance of the GATECDA method has improved compared to the five-fold validation, but its AUC (0.8876) and AUPR (0.8893) are still lower than that of the method proposed in this paper, thereby further demonstrating the superiority of the dual-mask contrastive learning framework we proposed in capturing complex biological association patterns.

TABLE 4 Performance comparison with the baseline under the 10-fold cross-validation.

Methods	AUC	AUPR	F1-score	Accuracy	Recall
SVM	0.8623	0.8533	0.8052	0.7977	0.8361
RF	0.8903	0.8906	0.8213	0.8176	0.8385
KNN	0.8700	0.8834	0.7970	0.7961	0.8005
AdaBoost	0.8836	0.8839	0.8240	0.8198	0.8439
VGAE	0.8634	0.8725	0.7987	0.7865	0.8314
VGAMF	0.8729	0.8682	0.8113	0.8030	0.8471
GCNMDA	0.8834	0.8864	0.8225	0.8183	0.8420
GATECDA	0.8876	0.8893	0.8237	0.8204	0.8442
MNGACDA	0.8901	0.8904	0.8210	0.8174	0.8383
Ours	0.8982	0.8973	0.8354	0.8329	0.8517

3.3 Ablation experiment

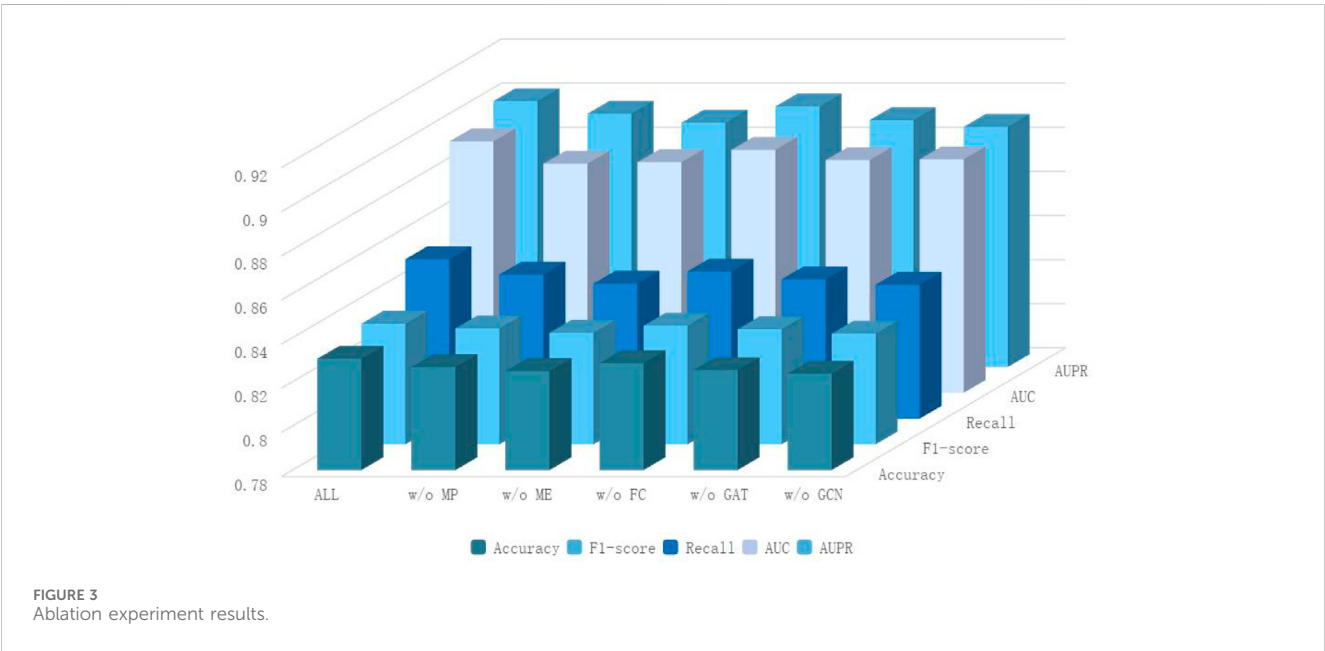
To dissect the contribution of each core component in our framework, we conducted a systematic ablation study, with results detailed in Table 5; Figure 3. We evaluated five ablated variants: removing the path mask (w/o MP), removing the edge mask (w/o ME), replacing the attention fusion classifier with an MLP (w/o FC), and removing either the GAT (w/o GAT) or GCN (w/o GCN) component from the encoder. The complete model achieves the best performance (AUC: 0.8940), confirming the synergistic design. The removal of either masking strategy (w/o MP or w/o ME) caused a substantial and nearly equivalent performance drop (AUC 0.884). This indicates that both path-level and edge-level masking are indispensable and complementary, each forcing the model to learn robust representations from different structural perspectives.

The complete model achieved the best performance (AUC: 0.8940), demonstrating the synergy among its components. Removing either the path mask (w/o MP) or edge mask (w/o ME) caused a noticeable performance drop (AUC: 0.884). This indicates that both masking strategies contribute uniquely to learning robust representations: path masking helps capture higher-order dependencies, while edge masking forces the model to rely on broader neighborhood structures. When the path masking module was removed, it was observed that all indicators showed significant declines, especially the AUC value dropping to 0.8839 and the recall rate dropping to 0.8453. This phenomenon indicates that the path-level structural information plays an important role in capturing the complex association patterns between circRNAs and drugs. The path masking effectively enhances the model's reasoning ability for graph structure dependencies by simulating the absence of local connectivity patterns. Similarly, the removal of the edge masking module also led to a significant performance reduction, with AUC dropping to 0.8846 and recall rate dropping to 0.8412. Edge masking weakens the direct connections between nodes at a finer granularity, prompting the model to pay more attention to other structural cues within the neighborhood. This strategy is crucial for learning robust graph representations. Notably, the performance decline of the path

TABLE 5 Comparison of ablation experiment results.

Methods/Metrics	AUC	AUPR	F1-score	Accuracy	Recall
ALL	0.8940	0.9006	0.8347	0.8305	0.8523
W/o MP	0.8839	0.8949	0.8326	0.8267	0.8453
W/o ME	0.8846	0.8908	0.8306	0.8249	0.8412
W/o FC	0.8902	0.8982	0.8338	0.8285	0.8467
W/o GAT	0.8855	0.8920	0.8322	0.8253	0.8431
W/o GCN	0.8859	0.8889	0.8302	0.8237	0.8407

Bold values indicate the best performance in each metric.



masking and edge masking modules was similar but with different focuses, indicating that the two masking strategies capture graph structure information at different levels, and their combined use can produce a complementary effect, jointly enhancing the model’s representation learning ability.

In terms of the feature fusion mechanism, when we replaced the attention-based feature fusion classifier with a simple multi-layer perceptron, the performance showed a slight but consistent decline, with AUC dropping to 0.8902. This result indicates that the attention mechanism can effectively allocate the importance weights of different features when integrating multi-view features, thereby improving the effect of feature fusion. However, the relatively small decline in performance also suggests that the MLP can still capture the interaction relationships between features to a certain extent, but lacks the ability to focus on important features prominently. The ablation experiments on the graph neural network components further revealed the complementary characteristics of GCN and GAT. When only using GCN and removing GAT, AUC dropped to 0.8855; while

only using GAT and removing GCN, AUC dropped to 0.8859. GCN captures the local topological patterns of the graph through the aggregation of neighbor node features, while GAT dynamically emphasizes important structural features through attention weights. The combination of the two enables the model to fully utilize the regular structure of the graph and adaptively focus on key information.

3.4 Case experiment

In this study, representative anticancer drugs (Doxorubicin, Gefitinib, Sorafenib, and Paclitaxel) from the CTRP (Cancer Therapeutics Response Portal) database were selected as an independent validation set to test the model’s generalization capability and biological interpretability in real biological scenarios. Numerous existing works have utilized public drug response databases (such as CTRP, GDSC, CCLE) to validate the extrapolation performance of drug–gene or drug–RNA prediction

TABLE 6 The top 20 circRNA-drug sensitivity association prediction results of doxorubicin.

Ranking	circRNA	Verification	Ranking	circRNA	Verification
1	HNRNPA2B1	✓	11	COL4A2	✓
2	CALD1	✓	12	FBLN1	×
3	VIM	✓	13	LTBP3	✓
4	COL3A1	✓	14	KRT7	✓
5	LINC01089	×	15	CTTN	✓
6	DCN	×	16	COL6A2	✓
7	HMGA2	✓	17	TGFB1	✓
8	ANXA2	✓	18	PKM	×
9	FN1	✓	19	MGAT4B	✓
10	EFEMP1	✓	20	COL8A1	×

TABLE 7 The top 20 circRNA-drug sensitivity association prediction results of gefitinib.

Ranking	circRNA	Verification	Ranking	circRNA	Verification
1	VIM	✓	11	FBLN1	✓
2	CALD1	✓	12	LINC01089	×
3	COL6A2	✓	13	KRT7	✓
4	HSP90B1	✓	14	CTTN	✓
5	FN1	✓	15	TGFB1	✓
6	COL4A1	✓	16	DCN	×
7	ANXA2	✓	17	COL8A1	×
8	PKM	✓	18	HNRNPA2B1	✓
9	MGAT4B	✓	19	LTBP3	✓
10	EFEMP2	×	20	HMGA2	×

models. Compared to these studies, the differences of this work lie in: (1) We predict circRNA–drug sensitivity associations, which represent a structured yet relatively sparse and noisy biological network problem; (2) The model employs multi-view encoding and various self-supervised enhancement strategies to improve the representation learning capability for complex graph structures, thereby aiming to maintain prediction accuracy even for unseen drugs/samples. By testing the model’s top 20 candidate circRNAs predicted on CTRP and comparing them with known experimental or literature-validated results in the database, we can directly compare our method with existing sequence/network-based or single GNN-based methods: if the validation rate on the independent database is significantly higher than random or baseline models, it indicates that our method holds an advantage in capturing biologically relevant signals, thereby establishing a connection with and surpassing relevant prior works. These drugs include doxorubicin, gefitinib, sorafenib, and paclitaxel, which represent different mechanisms of action of chemotherapy drugs and have good clinical representativeness and biological diversity. The experiment sorted the circRNA-drug sensitivity

associations by prediction scores, selected the top 20 high-confidence associations in the CTRP database for independent external validation, and used the validation rate as the main evaluation indicator.

In the validation results of doxorubicin (Table 6), 15 out of the top 20 predicted circRNAs were validated, with a validation rate of 75%. It is noteworthy that the top 4 circRNAs, including HNRNPA2B1, CALD1, VIM, and COL3A1, were all validated as significantly associated, indicating that the model has a good ability to identify highly confident associations. Doxorubicin, as an anthracycline broad-spectrum anti-tumor drug, induces DNA damage by integrating into the DNA double helix and activates the apoptosis signaling pathway. These validated circRNAs may play an important regulatory role in its mechanism of action.

Gefitinib, as a selective EGFR tyrosine kinase inhibitor, is mainly used for the treatment of EGFR-mutated non-small cell lung cancer. According to the verification results (Table 7), 15 out of the top 20 predicted associations were verified, with a verification rate of 75%. It is worth noting that EFEMP2 ranked 10th was not verified, while HNRNPA2B1 ranked 18th was verified. This phenomenon

TABLE 8 The top 20 circRNA-drug sensitivity association prediction results of sorafenib.

Ranking	circRNA	Verification	Ranking	circRNA	Verification
1	COL3A1	✓	11	HSP90B1	✓
2	VIM	✓	12	FN1	✓
3	ANXA2	✓	13	KRT7	✓
4	CALD1	✓	14	LTBP3	✓
5	COL6A2	✓	15	CTTN	✓
6	FBLN1	✓	16	DCBLD2	✓
7	PKM	✓	17	HMGA2	×
8	LINC01089	×	18	MGAT4B	✓
9	HNRNPA2B1	✓	19	COL8A1	×
10	EFEMP1	✓	20	COL4A2	×

TABLE 9 The top 20 circRNA-drug sensitivity association prediction results of paclitaxel.

Ranking	circRNA	Verification	Ranking	circRNA	Verification
1	ANXA2	✓	11	FBLN1	✓
2	CALD1	✓	12	KRT7	✓
3	VIM	✓	13	COL6A2	✓
4	FN1	✓	14	PKM	✓
5	HSP90B1	✓	15	LINC01089	×
6	DCN	×	16	CTTN	✓
7	HMGA2	✓	17	LTBP3	✓
8	COL3A1	✓	18	MGAT4B	✓
9	EFEMP1	✓	19	TGFB1	✓
10	COL4A2	✓	20	HNRNPA2B1	✓

indicates that the model still has room for optimization in the discrimination of certain moderate confidence associations, possibly related to the expression heterogeneity of circRNA in specific cell lines.

In the validation analysis of sorafenib (Table 8), the model demonstrated superior predictive performance, with 16 out of 20 predicted associations confirmed, yielding a validation rate of 80%. As a multi-target tyrosine kinase inhibitor, sorafenib is widely used in the treatment of hepatocellular carcinoma and renal cell carcinoma. Notably, all of the top six ranked circRNAs—including COL3A1, VIM, and ANXA2—were experimentally validated, indicating high prediction accuracy for top-ranking associations. These circRNAs may modulate tumor cell sensitivity to sorafenib through the regulation of multiple signaling pathways.

The most encouraging result comes from the validation analysis of paclitaxel (Table 9). Among the first 20 predicted correlations, 18 were validated, with a validation rate of up to 90%. Paclitaxel, as a microtubule stabilizer, inhibits microtubule depolymerization to block the cell cycle and is widely used in the treatment of breast

cancer and ovarian cancer. This excellent predictive performance indicates that the dual-mask graph contrastive learning framework proposed in this paper has a special advantage for the prediction of sensitivity correlations of microtubule-targeted drugs. From the validation results, it can be seen that multiple cytoskeleton-related circRNAs including ANXA2, CALD1, and VIM have been validated, which is highly consistent with the mechanism of paclitaxel. Through the comprehensive analysis of the validation results of the four drugs (Table 10), the average validation rate of this model reached 80%, demonstrating good generalization ability. Further analysis revealed that some circRNAs were predicted to have high-ranking correlations in multiple drugs and were validated, such as VIM, CALD1, ANXA2, and FN1, etc. These circRNAs may be involved in the universal drug resistance mechanism of tumor cells, and their molecular functions are worthy of in-depth study through subsequent experiments. Additionally, the model's predictions for some circRNAs (such as LINC01089, COL8A1) showed deviations, which may be related to the incomplete annotation or low expression level of these circRNAs in the CTRP database.

TABLE 10 Summary of the validation rates of the top 20 predicted associations for the four drugs.

Drug	Verified association count	Total number of associations	Verification rate
Doxorubicin	15	20	75%
Gefitinib	15	20	75%
Sorafenib	16	20	80%
Paclitaxel	18	20	90%
Average	16	20	80%

TABLE 11 The impact of the number of GCN and GAT layers on performance.

GCN and GAT layers	AUC	AUPR
1	0.8911	0.8957
2	0.8922	0.8963
3	0.8940	0.9006
4	0.8914	0.8953
5	0.8897	0.8944

TABLE 12 Comparison of model performance metrics.

Learning rate	AUC	AUPR
0.1	0.8765	0.8691
0.01	0.8864	0.8882
0.001	0.8940	0.9006
0.0001	0.8875	0.8927
0.00001	0.8868	0.8915

3.5 Parameter experiment

To determine the optimal parameter configuration of the model and to deeply understand the influence mechanism of each hyperparameter on the model's performance, we conducted further parameter experiments on three key parameters: the number of GCN and GAT layers, the hidden layer dimension of the multi-view graph encoder, and the Dropout rate. These parameters respectively affect the model's expression ability and generalization performance from three different dimensions: model depth, representation ability, and regularization strength. All experiments were conducted under the setting of five-fold cross-validation, with AUC and AUPR as the main evaluation indicators to ensure the statistical reliability of the results. Table 11 shows the impact of the number of GCN and GAT layers on the model's performance. As the number of network layers increased from 1 layer to 3 layers, the model's AUC improved from 0.8911 to 0.8940, and AUPR improved from 0.8957 to 0.9006. This indicates that appropriately increasing the network depth helps the model capture more complex graph structure features. However, when the number of layers continued to increase to 4 layers and 5 layers, the performance slightly decreased, with AUC dropping to 0.8914 and 0.8897 respectively. This might be due to the overfitting or gradient vanishing problem caused by an overly deep network. This phenomenon is consistent with the theoretical analysis of graph neural networks, that is, an excessively deep GNN layer may lead to the oversmoothing of node features, thereby reducing the model's discriminative ability. Finally, 3 layers were selected as the optimal configuration for GCN and GAT, achieving a good balance between model expression ability and training stability.

To determine the optimal learning rate for model training, we conducted a sensitivity analysis by evaluating a spectrum of values ranging from 0.1 to 0.00001. The performance, measured by AUC

and AUPR, is summarized in Table 12. The results demonstrate that the learning rate significantly impacts model efficacy. A relatively high learning rate of 0.1 led to suboptimal performance (AUC = 0.8765, AUPR = 0.8691), suggesting potential instability during the gradient descent process. As the learning rate decreased to 0.01, the performance improved notably. The best performance was achieved with a learning rate of 0.001, yielding the highest AUC (0.8940) and AUPR (0.9006). Further decreasing the learning rate to 0.0001 and 0.00001 resulted in a slight but consistent performance degradation, indicating that overly small learning rates might hinder the model's convergence to an optimal solution. Consequently, a learning rate of 0.001 was selected as the optimal configuration for all subsequent experiments, striking a balance between training efficiency and model performance.

In terms of the selection of the hidden layer dimensions in the multi-view graph encoder, we compared different dimension settings ranging from 16 to 256. The experimental results are shown in Table 13. When the hidden layer dimension was set to 32, the model achieved the best performance, with an AUC of 0.8940 and an AUPR of 0.9006. Smaller dimension settings (such as 16) might limit the model's representational ability, resulting in an AUC of only 0.8882; while overly large dimensions (such as 64, 128, 256) might introduce too many parameters, increasing the risk of overfitting and causing performance to decline to varying degrees. This result indicates that an appropriate hidden layer dimension can achieve the best balance between model capacity and generalization ability, ensuring sufficient feature representation space while avoiding optimization difficulties caused by parameter redundancy.

The Dropout rate, as a key parameter for controlling the regularization intensity of the model, has an impact as shown in Table 14. When the Dropout rate is set to 0.3, the model exhibits the best performance, with an AUC of 0.8940 and an AUPR of 0.9006. A lower Dropout rate (0.1) may lead to insufficient regularization,

TABLE 13 The impact of hidden layers in multi-view graph encoders on performance.

Hidden layers	AUC	AUPR
16	0.8882	0.8943
32	0.8940	0.9006
64	0.8864	0.8938
128	0.8907	0.8983
256	0.8857	0.8924

Bold values indicate the best performance in each metric.

TABLE 14 The impact of dropout rate on performance.

Dropout rate	AUC	AUPR
0.1	0.8907	0.8989
0.3	0.8940	0.9006
0.5	0.8859	0.8936
0.7	0.8671	0.8783

Bold values indicate the best performance in each metric.

causing the model to overfit the training data; while a higher Dropout rate (0.5 and 0.7) may overly suppress the neuron activation, resulting in underfitting of the model, especially when the Dropout rate reaches 0.7, the performance significantly drops to an AUC of 0.8671 and an AUPR of 0.8783. This trend clearly demonstrates the dual role of the Dropout mechanism in deep learning models: an appropriate Dropout rate can effectively prevent overfitting and enhance the model’s generalization ability; but an overly strong Dropout rate will damage the model’s learning ability and lead to performance degradation.

Based on the results of the parameter experiments, the optimal parameter configuration of the model was determined: the number of layers for GCN and GAT is 3, the hidden layer dimension is 32, and the Dropout rate is 0.3. This configuration achieved the best performance in all three key parameters. Through this experiment, not only did it provide a reliable parameter basis for the method in this paper, but it also offered valuable references for other bioinformatics tasks based on graph neural networks.

4 Conclusion

This study introduces a dual-mask graph contrastive learning (DMAGCL) framework for predicting circRNA-drug sensitivity associations. The core innovation lies in its synergistic use of path- and edge-level masking to learn robust node representations, an adaptive contrastive loss to dynamically balance exploration and exploitation during training, and an attention fusion classifier to effectively integrate multi-modal features. Extensive evaluations confirmed that DMAGCL achieves state-of-the-art performance, with case studies on anti-cancer drugs demonstrating its biological relevance and an average verification rate of 80. Looking forward, this work opens several promising directions: First, the framework could be applied to related tasks like miRNA-disease association prediction,

specifically to model the complex interactions in competing endogenous RNA (ceRNA) networks. Second, integrating multi-omics data (e.g., transcriptomics from specific cell lines) could help address the heterogeneity observed in some predictions and improve cell context-specific modeling. Finally, exploring the integration of large language models for biological sequence encoding could further enrich feature representations. This study thus establishes a robust and extensible computational paradigm with significant potential for biomarker discovery and drug sensitivity analysis in precision oncology.

Data availability statement

The data and code presented in the study are publicly available at <https://github.com/yuqi23-guo/DMAGCL>.

Author contributions

PW: Methodology, Data curation, Supervision, Funding acquisition, Resources, Writing – review and editing, Conceptualization, Writing – original draft. YG: Writing – review and editing, Formal Analysis, Software, Data curation. ZL: Writing – original draft, Visualization, Software. DT: Software, Visualization, Writing – original draft. MQ: Investigation, Validation, Writing – review and editing, Supervision. ZZ: Writing – original draft, Investigation, Validation. LZ: Writing – review and editing, Resources.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This research was financially supported by the National Natural Science Foundation of China (No. 62172158), the Natural science foundation of Fujian Province (Grant No.2024J08372) and the National Natural Science Foundation of China (No. 62471318).

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.

The reviewer WL declared a past co-authorship with the author ZL to the handling editor.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Ai, N., Yuan, H., Liang, Y., Lu, S., Ouyang, D., Lai, Q. H., et al. (2024). Multi-view multitattention graph learning with stack deep matrix factorization for circrna-drug sensitivity association identification. *IEEE J. Biomed. Health Inf.* 28, 7670–7682. doi:10.1109/jbhi.2024.3431693
- Baptista, D., Correia, J., Pereira, B., and Rocha, M. (2022). Evaluating molecular representations in machine learning models for drug response prediction and interpretability. *J. Integr. Bioinforma.* 19, 20220006. doi:10.1515/jib-2022-0006
- Bartel, D. P. (2004). Micrnas: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi:10.1016/s0092-8674(04)00045-5
- Cai, L., Yue, G., Chen, Y., Wang, L., Yao, X., Zou, Q., et al. (2025). Et-protacs: modeling ternary complex interactions using cross-modal learning and ternary attention for accurate protac-induced degradation prediction. *Briefings Bioinforma.* 26, bbac654. doi:10.1093/bib/bbae654
- Chen, L.-L., and Yang, L. (2015). Regulation of circrna biogenesis. *RNA Biol.* 12, 381–388. doi:10.1080/15476286.2015.1020271
- Chen, Y., Wang, J., Wang, C., Liu, M., and Zou, Q. (2022). Deep learning models for disease-associated circrna prediction: a review. *Briefings Bioinforma.* 23, bbac364. doi:10.1093/bib/bbac364
- Deng, L., Liu, Z., Qian, Y., and Zhang, J. (2022). Predicting circrna-drug sensitivity associations via graph attention auto-encoder. *BMC Bioinforma.* 23, 160. doi:10.1186/s12859-022-04694-y
- Ding, C., Yi, X., Chen, X., Wu, Z., and You, H. (2021). Warburg effect-promoted exosomal circ_0072083 releasing up-regulates NANO expression through multiple pathways and enhances temozolomide resistance in glioma. *J. Exp. and Clin. Cancer Res.* 40, 164. doi:10.1186/s13046-021-01942-6
- Du, X., and Xue, Z. (2022). Jlcrcb: a unified multi-view-based joint representation learning for circrna binding sites prediction. *J. Biomed. Inf.* 136, 104231. doi:10.1016/j.jbi.2022.104231
- Fan, Z., Zhang, Y., Li, Y., Zhong, Z., and Deng, L. (2024). "Lnsncda: unraveling circrna-drug sensitivity via local smoothing graph neural network and credible negative samples," in 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
- Fu, X., Du, Z., Chen, Y., Chen, H., Zhuo, L., Lu, A., et al. (2025a). Drugkans: a paradigm to enhance drug-target interaction prediction with Kans. *IEEE J. Biomed. Health Inf.* 1–12. doi:10.1109/JBHI.2025.3566931
- Fu, X., Peng, L., Chen, H., Rong, M., Chen, Y., Cao, D., et al. (2025b). Grape: graph-regularized protein language modeling unlocks tcr-epitope binding specificity. *Briefings Bioinforma.* 26, bbaf522. doi:10.1093/bib/bbaf522
- Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57–70. doi:10.1016/s0092-8674(00)81683-9
- Lan, W., Dong, Y., Zhang, H., Li, C., Chen, Q., Liu, J., et al. (2023). Benchmarking of computational methods for predicting circrna-disease associations. *Briefings Bioinforma.* 24, bbac613. doi:10.1093/bib/bbac613
- Li, X., Yang, L., and Chen, L.-L. (2018). The biogenesis, functions, and challenges of circular rnas. *Mol. Cell* 71, 428–442. doi:10.1016/j.molcel.2018.06.034
- Li, G., Li, Y., Liang, C., and Luo, J. (2023). Deepwalk-aware graph attention networks with cnn for circrna-drug sensitivity association identification. *Briefings Funct. Genomics* 23, 418–428. doi:10.1093/bfpg/eld053
- Li, M., Coşkun, M., and Koyutürk, M. (2025a). Topological-similarity based canonical representations for biological link prediction. *IEEE Trans. Comput. Biol. Bioinforma.* 22, 1278–1287. doi:10.1109/tcb.2024.3462730
- Li, X., Wang, J., and Yan, Z. (2025b). Can graph neural networks be adequately explained? A survey. *ACM Comput. Surv.* 57, 1–36. doi:10.1145/3711122
- Liu, Y., Dong, Y., Zhao, L., Su, L., and Luo, J. (2018). Circular rna-mto1 suppresses breast cancer cell viability and reverses monastrol resistance through regulating the traf4/eg5 axis. *Int. J. Oncol.* 53, 1752–1762. doi:10.3892/ijo.2018.4485
- Liu, H., Chen, C., Su, Y., Zuo, E., Wu, L., Li, M., et al. (2024). Tensor product graph diffusion based on nonlinear fusion of multi-source information to predict circrna-disease associations. *Appl. Soft Comput.* 152, 111215. doi:10.1016/j.asoc.2023.111215
- Liu, T., Wang, S., Zhang, Y., Pang, S., Yin, W., Wu, W., et al. (2025a). Deciphering circrna-drug sensitivity associations via global-local heterogeneous matrix factorization and hypergraph contrastive learning. *Expert Syst. Appl.* 292, 128548. doi:10.1016/j.eswa.2025.128548
- Liu, Z., Dai, Q., Yu, X., Duan, X., and Wang, C. (2025b). Predicting circrna-drug resistance associations based on a multimodal graph representation learning framework. *IEEE J. Biomed. Health Inf.* 29, 1838–1848. doi:10.1109/jbhi.2023.3299423
- Lu, S., Liang, Y., Li, L., Liao, S., Zou, Y., Yang, C., et al. (2023). Inferring circrna-drug sensitivity associations via dual hierarchical attention networks and multiple kernel fusion. *BMC Genomics* 24, 796. doi:10.1186/s12864-023-09899-w
- Ma, J., Qin, T., Zhai, M., and Cai, L. (2024). Agcnaf: predicting disease-gene associations using gcn and multi-head attention to fuse the similarity features. *Eng. Res. Express* 6, 045221. doi:10.1088/2631-8695/ad8c9f
- Milano, M., and Cannataro, M. (2023). Network models in bioinformatics: modeling and analysis for complex diseases. *Briefings Bioinforma.* 24, bbad016. doi:10.1093/bib/bbad016
- Min, S., Xiao, Y., and Ma, J. (2019). Circular rnas in cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Mol. Cancer* 18, 1–17. doi:10.1186/s12943-019-1002-6
- Niu, M., Wang, C., Chen, Y., Zou, Q., and Luo, X. (2025). Interpretable multi-instance heterogeneous graph network learning modelling circrna-drug sensitivity association prediction. *BMC Biol.* 23, 131. doi:10.1186/s12915-025-02223-w
- Peng, L., Li, H., Yuan, S., Meng, T., Chen, Y., Fu, X., et al. (2025a). Metacda: a novel framework for circrna-driven drug discovery utilizing adaptive aggregation and meta-knowledge learning. *J. Chem. Inf. Model.* 65, 2129–2144. doi:10.1021/acs.jcim.4c02193
- Peng, L., Wang, W., Yang, Z., Fu, X., Liang, W., and Cao, D. (2025b). Leveraging explainable multi-scale features for fine-grained circrna-mirna interaction prediction. *BMC Biol.* 23, 121. doi:10.1186/s12915-025-02227-6
- Ru, X., Ye, X., Sakurai, T., and Zou, Q. (2021a). Application of learning to rank in bioinformatics tasks. *Briefings Bioinforma.* 22, bbac394–11. doi:10.1093/bib/bbac394
- Ru, X., Ye, X., Sakurai, T., Zou, Q., Xu, L., and Lin, C. (2021b). Current status and future prospects of drug-target interaction prediction. *Briefings Funct. Genomics* 20, 312–322. doi:10.1093/bfpg/elab031
- Ru, X., Ye, X., Sakurai, T., and Zou, Q. (2022). Nerltr-dta: drug-target binding affinity prediction based on neighbor relationship and learning to rank. *Bioinformatics* 38, 1964–1971. doi:10.1093/bioinformatics/btac048
- Song, J., Wei, M., Zhao, S., Zhai, H., Dai, Q., and Duan, X. (2024). Drug sensitivity prediction based on multi-stage multi-modal drug representation learning. *Interdiscip. Sci. Comput. Life Sci.* 17, 231–243. doi:10.1007/s12539-024-00668-1
- Suzuki, H., and Tsukahara, T. (2014). A view of pre-mrna splicing from rna-seq resistant rnas. *Int. J. Mol. Sci.* 15, 9331–9342. doi:10.3390/ijms15069331
- Toprak, A. (2024). circrna-disease association prediction with an improved unbalanced bi-random walk. *J. Radiat. Res. Appl. Sci.* 17, 100858. doi:10.1016/j.jrras.2024.100858
- Turgut, H., Turanlı, B., and Boz, B. (2023). Dcda: circrna-disease association prediction with feed-forward neural network and deep autoencoder. *Interdiscip. Sci. Comput. Life Sci.* 16, 91–103. doi:10.1007/s12539-023-00590-y
- Wang, X., Wang, H., Jiang, H., Qiao, L., and Guo, C. (2023). Circular rna circ_0076305 promotes cisplatin (ddp) resistance of non-small cell lung cancer cells by regulating abcc1 through mir-186-5p. *Cancer Biotherapy Radiopharm.* 38, 293–304. doi:10.1089/cbr.2020.4153
- Wang, Y., Lei, X., Chen, Y., Guo, L., and Wu, F.-X. (2025a). Circular rna-drug association prediction based on multi-scale convolutional neural networks and adversarial autoencoders. *Int. J. Mol. Sci.* 26, 1509. doi:10.3390/ijms26041509
- Wang, Z., Lei, X., Zhang, Y., Wu, F.-X., and Pan, Y. (2025b). Recent progress of deep learning methods for rbp binding sites prediction on circrna. *Curr. Bioinforma.* 20, 487–505. doi:10.2174/0115748936308564240712053215
- Wei, J., Zhuo, L., Pan, S., Lian, X., Yao, X., and Fu, X. (2023a). Headtailtransfer: an efficient sampling method to improve the performance of graph neural network method in predicting sparse ncna-protein interactions. *Comput. Biol. Med.* 157, 106783. doi:10.1016/j.cbiomed.2023.106783
- Wei, J., Zhuo, L., Zhou, Z., Lian, X., Fu, X., and Yao, X. (2023b). Gcfmcl: predicting mirna-drug sensitivity using graph collaborative filtering and multi-view contrastive learning. *Briefings Bioinforma.* 24, bbad247. doi:10.1093/bib/bbad247
- Xiao, Q., Dai, J., and Luo, J. (2021a). A survey of circular rnas in complex diseases: databases, tools and computational methods. *Briefings Bioinforma.* 23, bbab444. doi:10.1093/bib/bbab444

Xiao, Q., Fu, Y., Yang, Y., Dai, J., and Luo, J. (2021b). Nsl2cd: identifying potential circrna-disease associations based on network embedding and subspace learning. *Briefings Bioinforma.* 22, bbab177. doi:10.1093/bib/bbab177

Xiao, X., Qian, Y., Huang, Z., Zheng, R., and Deng, L. (2023). "Predicting associations between circrnas and drug sensitivity using heterogeneous graphs and graph attention networks," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 191–196. doi:10.1109/bibm58861.2023.10385513

Xie, F., Xiao, X., Tao, D., Huang, C., Wang, L., Liu, F., et al. (2020). circnr3c1 suppresses bladder cancer progression through acting as an endogenous blocker of brd4/c-myc complex. *Mol. Ther. - Nucleic Acids* 22, 510–519. doi:10.1016/j.omtn.2020.09.016

Yang, B., and Chen, H. (2023). Predicting circrna-drug sensitivity associations by learning multimodal networks using graph auto-encoders and attention mechanism. *Briefings Bioinforma.* 24, bbac596. doi:10.1093/bib/bbac596

Yin, S., Xu, P., Jiang, Y., Yang, X., Lin, Y., Zheng, M., et al. (2024). Predicting the potential associations between circrna and drug sensitivity using a multisource feature-based approach. *J. Cell. Mol. Med.* 28, e18591. doi:10.1111/jcmm.18591

Zhang, P., Wang, Z., Sun, W., Xu, J., Zhang, W., Wu, K., et al. (2023). Rdrge: a framework for noncoding rna-drug resistance discovery by incorporating graph skeleton extraction and attentional feature fusion. *ACS Omega* 8, 27386–27397. doi:10.1021/acsomega.3c02763

Zhou, Z., Du, Z., Wei, J., Zhuo, L., Pan, S., Fu, X., et al. (2023). Mham-npi: predicting ncRNA-protein interactions based on multi-head attention mechanism. *Comput. Biol. Med.* 163, 107143. doi:10.1016/j.combiomed.2023.107143

Zhou, Z., Du, Z., Jiang, X., Zhuo, L., Xu, Y., Fu, X., et al. (2024a). Gam-mdr: probing mirna-drug resistance using a graph autoencoder based on random path masking. *Briefings Funct. Genomics* 23, 475–483. doi:10.1093/bfpg/ela005

Zhou, Z., Liao, Q., Wei, J., Zhuo, L., Wu, X., Fu, X., et al. (2024b). Revisiting drug-protein interaction prediction: a novel global-local perspective. *Bioinformatics* 40, btac271. doi:10.1093/bioinformatics/btac271

Zhuo, L., Song, B., Liu, Y., Li, Z., and Fu, X. (2022). Predicting ncRNA-protein interactions based on dual graph convolutional network and pairwise learning. *Briefings Bioinforma.* 23, bbac339. doi:10.1093/bib/bbac339