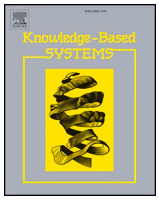
Knowledge-Based Systems 330 (2025) 114548



Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Heterogeneous graph collaborative representation learning for drug-related microbe prediction with attentive fusion and reciprocal distillation

Yanbu Guo [a], Quanming Guo [a], Shengli Song [a,*], Yihan Wang [a], Jinde Cao [b,c,**]

[a] College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China
[b] School of Mathematics, Southeast University, Nanjing 211189, China
[c] Purple Mountain Laboratories, Nanjing 211111, China

ARTICLE INFO

Keywords:
Complex networks
Graph transformer
Representation learning
Drug-microbe associations

ABSTRACT

Microbes are microorganisms with biological molecules and have significant therapeutic potential for treating diseases, underscoring the need for computational methods to screen microbes targeting disease-associated drugs. However, the computational methods often consider node embedding or structure features between microbes and drugs, and have a severe class imbalance problem inherent in sparse association data. In this work, we proposed a heterogeneous graph collaborative representation learning model that combines the merits of attentive fusion and reciprocal distillation for drug-related microbe prediction. First, we constructed the heterogeneous biological information and meta-path-induced graphs of microbes and drugs. Then, a topological structure feature encoder is devised to extract complex topological and semantic interaction patterns from heterogeneous biological graphs with microbes and drugs, while an efficient transformer concurrently extracts discriminative semantic and structural information based on the graph position information of nodes. Next, a reciprocal distillation schema is developed to mitigate the adverse effects of the data imbalance problem, and enable the distribution consistency of the model between topological and semantic information extraction. Moreover, we devised a dual collaborative feature fusion schema that combines graph topological and dual meta-path-based semantic features to obtain the discriminative features of microbes and drugs. Through reciprocal distillation, an efficient optimization function focuses on hard-to-classify samples of drug-related microbes via discriminative features. Extensive experiments demonstrate that our model could deal with the association sparsity problem and extract more semantics and structure. Meanwhile, case studies indicate that our model could discover reliable candidate microbes associated with a special drug.

1. Introduction

Microbe communities are composed of bacteria, archaea, and eukaryotes, and often produce essential vitamins, facilitate metabolism, and protect against invasion from pathogens. The occurrence of diseases [1] is closely related to the microbial community of the human body, and microorganisms influence tumorigenesis and metastasis [2]. The increasing studies of abnormal microbial responses to various drugs highlight the potential of the microbes as valuable diagnostic markers and therapeutic targets. However, discovering the drug-related microbes [3] is time-consuming and labour-intensive through biological wet experiments, and currently, computational methods [4] often consider node embedding or structure features between microbes and drugs. Consequently, exploring efficient methods is an urgent need to discover the potential association of microbes and drugs for academia and industry.

In recent years, researchers have developed various computational methods to screen drug-microbe association information, which mainly includes microbe and drug-based computational methods [5–8] and microbe, disease, and drug-based computational methods [9–12]. The main differences of the above methods are that microbe, disease, and drug-based methods consider disease data as additional contextual features for accurately mining microbe-drug association information. Besides, current drug-associated microbe prediction methods [13–15] fail to sufficiently exploit the topology and position information, and rely solely on association matrices to extract features of drugs and

* Corresponding author.
** Corresponding author at: School of Mathematics, Southeast University, Nanjing 211189, China.
E-mail addresses: 2003010@zzuli.edu.cn (S. Song), jdcao@seu.edu.cn (J. Cao).

https://doi.org/10.1016/j.knosys.2025.114548
Received 28 July 2025; Received in revised form 22 September 2025; Accepted 23 September 2025
Available online 25 September 2025
0950-7051/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

microbes, and known microbe-drug associations are reliant on microbe and drug similarities. In practice, a significant amount of drug-related data has heterogeneity and long-distance correlations of multiple types of nodes and connections, and different node features have unequal dimensions and scales, and each node has a distinct structural importance. Although these methods have demonstrated remarkable performance for drug-microbe association prediction, there is still room to enhance the association mining performance for the features of microbes and drugs. Inspired by the successful applications [16,17], we injected some information about the relative or absolute position of the nodes in the graph to make use of the node order of the graph.

Graph neural networks have been applied into various fields such as polyadenylation signals [18,19], translation initiation sites [20] and miRNA-disease associations [21], and graph contrastive learning [8,13] has shown great advantages for extracting important features of drugs, microbe and drug nodes, but it fails to fully capture the complex structural information of heterogeneous biological graphs. Moreover, the performance of the contrastive learning-based models is often hampered by the severe data sparsity and class imbalance from known datasets of drug-microbe associations. Thus, how to select truly informative negative samples [13] remains a significant challenge in drug-related microbe prediction. The prediction efficacy of drug-related microbes crucially relies on the biological graph construction of positive and negative sample pairs [8,13,22], and the unobserved interactions of microbes and drugs are often treated as negative samples for microbe-drug association prediction. However, such unobserved drug-microbe pairs often reflect a lack of experimental investigation in sparse biological datasets rather than true non-interactions between microbes and drugs, leading to false negatives that can mislead training [22].

Besides, recent advances have shown promise in other biological domains such as Transformer-based models [13,23,24] and knowledge distillation techniques [25,26]. However, the direct application to heterogeneous drug-microbe graphs is challenging. Transformer-based models, originally designed for sequential data like text, have difficulty effectively capturing the complex, non-Euclidean topology of biological networks. Their standard positional embeddings do not adequately represent the relative structural importance of nodes, and their uniform attention mechanisms do not adapt to the different scales of information (from local neighbor nodes to global structures) captured across graph layers. Similarly, while knowledge distillation can improve model robustness, existing methods primarily focus on teacher-student frameworks and have not been tailored to address the unique challenges of drug-microbe prediction, such as high data sparsity, the need for a framework that learns without reliable negative samples, and the fundamental challenge of learning from both topological and semantic information simultaneously.

To address the above limitations for drug-related microbe prediction, we devised a heterogeneous graph collaborative representation learning with reciprocal distillation called HGCRL for extracting structural and topological biological patterns. The model primarily represents a compositional innovation, drawing upon and adapting established techniques such as graph transformers, meta-path-based convolutions, and distillation strategies from related fields like natural language processing and computer vision. At the same time, the proposed model incorporates elements of novel design principles through specialized components, such as the layer-adaptive mechanisms and reciprocal pathways tailored to handle microbial and drug heterogeneity in sparse association scenarios. Specifically, HGCRL mainly includes heterogeneous graph construction, topological structure feature learning, meta-path-based semantic representation learning, and an association predictor. We first constructed the heterogeneous biological graphs of microbes and drugs, as well as meta-path-induced graphs. Then, HGCRL takes representations of microbes and drugs extracted from meta-path-induced graphs as input and extracts the discriminative representations via a dual collaborative feature learning schema. Moreover, the features of microbes and drugs are enhanced with the reciprocal distillation strategy to obtain discriminative representations from heterogeneous similarity graphs. The reciprocal distillation module provides a dual learning pathway to extract robust representations without reliance on explicit negative sampling. Extensive experiments demonstrate the effectiveness of the proposed model to extract discriminative features and suppress irrelevant features of microbe-drug associations. The main contributions are summarized as follows:

1) We proposed a deep drug-related microbe prediction model that combines a layer-adaptive graph transformer with reciprocal distillation in a graph collaborative representation learning way and subgraph positional encoding. The model could deal with the graph heterogeneity of drugs and microbes through node and edge-type-dependent representations, and extract topological and semantic features that include position and topological features of microbe and drug nodes.
2) The layer-adaptive graph transformer is devised to extract topological and semantic patterns of microbes and drugs based on position-aware representations, which combine scaled dot-product attentive fusion, and node information aggregation. The meta-path-based semantic feature encoder is devised to extract the semantic representation of microbes and drugs based on two different meta-paths and graph convolutions
3) The reciprocal distillation module is devised to enhance graph structural and topological information of microbe and drug nodes through attentive fusion. Instead of contrasting sample pairs for drug-related microbe prediction, the special distillation module has two special paths to reciprocally distil representations of nodes, and further enforces the representation consistency of drug and microbe features.
4) The efficient optimization function is designed to mitigate the association sparsity of drug-microbe associations, which enhances robust structural and topological representation learning. The special function could effectively mitigate the adverse effects of data imbalance inherent in drug-microbe association prediction.

2. Related work

2.1. Computational methods-based association prediction

Microbe and drug-based computational methods have used information related only to microbes and drugs. The shallow features of microbes and drugs extracted by traditional algorithms struggle to extract biological patterns and intricate relationships of complex biological networks. For example, Huang et al. [7] developed the graph-normalized convolutional model for the potential microbe-drug association prediction. Wang et al. [27] proposed an integrated model with graph convolutions and attention algorithms to capture interaction patterns and discriminative features of nodes. The above methods achieved great results, but they ignored positive and negative sample imbalance and the isolated node problem. Then, Tian et al. [8] incorporated structure-enhanced contrastive learning with adaptive negative sampling strategies to predict a drug-related microbe prediction model. However, microbe and drug-based computational methods have used information related only to microbes and drugs. Then, the researchers have used disease information as additional nodes to explore the microbe-drug association prediction. For example, Long et al. [9] constructed a microbe-drug and microbe-disease-drug graph and proposed a graph attentive model for potential microbe-drug association prediction. Based on a heterogeneous graph of drugs, microbes, and diseases, Fan et al. [28] obtained the known microbe-disease-drug associations, and developed stacked autoencoders with attention and a random walk to learn attributes of nodes. Similarly, Wang et al. [29] extended these ideas to microbe-disease associations, proposing GCDSAEMDA that employs multi-head graph attention and deep sparse autoencoders, with

cosine distance-based k-means clustering for negative sample selection, achieving superior performance on HMDAD and Disbiome databases. After constructing heterogeneous graphs of drugs, microbes, and diseases, Li et al. [30] constructed a heterogeneity graph with drugs, microbes, and diseases, and developed a potential drug-microbe association prediction model based on matrix factorization and a three-layer heterogeneous biological graph.

2.2. Advanced network metrics and graph reasoning methods

Beyond traditional graph neural networks, recent research has explored efficient network metrics and advanced reasoning mechanisms. Ullah et al. [31] proposed OVED-Rank, which uses effective distance to replace computationally expensive shortest-path algorithms, combining node degree, k-shell index, and triangle count for identifying influential spreaders. Similarly, Ullah et al. [32] introduced electric potential centrality (EPC) that incorporates both local connections and global network structure to identify key influencers, demonstrating superior performance in the susceptible-infected-recovered model. Additionally, Li et al. [33] designed a remarkable heterogeneous graph reasoning model that assigns attention weights to relational subgraphs and employs bidirectional propagation to integrate semantic reasoning with large language models. Heterogeneous graph neural networks have emerged as powerful tools for modeling complex graph-structured data, and these advances in network metrics and reasoning mechanisms provide valuable insights for handling computational challenges in heterogeneous biological graphs. However, the hybrid models encounter feature space discrepancies during alignment and fusion, and suffer from over-smoothing problems with multi-layer convolutions.

2.3. Transformer-based and knowledge distillation methods

Recent advances in transformer and knowledge distillation have opened new avenues for biological network analysis. Transformer-based models have demonstrated superior capability in capturing long-range dependencies and complex patterns. For example, Zhang et al. [23] introduced metapath-based hierarchical transformers to model multi-scale interactions in drug-target networks. Qiu et al. [24] integrated graph isomorphism network, DenseNet, and BERT to create a multi-modal framework that processes both graph structures and sequence information. Li et al. [13] developed a pre-trained sequence-to-sequence model specifically for biological predictions, achieving state-of-the-art performance on adverse drug reaction tasks. Besides, knowledge distillation has emerged as a powerful technique for enhancing model performance without explicit negative sampling. For example, Xia et al. [25] employed a teacher–student framework to handle unseen proteins and compounds effectively. While these methods show promising results in their respective domains, they have not been specifically designed for drug–microbe association prediction. For example, Wang et al. [26] combined the federated knowledge distillation technique to accomplish synergistic compound-protein interaction prediction.

3. Materials and methods

As shown in Fig. 1, we devised the heterogeneous graph collaborative representation learning model for drug-related microbe prediction with attentive fusion and reciprocal distillation. HGCRL includes heterogeneous graph and meta-path construction, topological structure feature encoder, meta-path-based semantic feature encoder, and association prediction.

3.1. Datasets and similarity calculation

1) Statistics of datasets and similarity of drugs and microbes

To validate the effectiveness and robustness, the microbe-drug association database(MDAD) [34], a resource of anti-biofilm agents (aBiofilm) [35], and DrugVirus [36]datasets were employed for drug-related microbe prediction. The MDAD dataset contains 2,470 experimentally verified associations between 173 microbes and 1,373 drugs. The aBiofilm dataset comprises 2,884 microbe-drug associations involving 140 microbes and 1,720 drugs. The DrugVirus dataset consists of 933 microbe-drug associations between 95 microbes and 175 drugs. N_m and N_d denote the number of microbes and drugs, respectively, and the microbe-drug associations are represented by an adjacency matrix $A \in \mathbb{R}^{N_m \times N_d}$. $A_{i,j}$ is the value '1' when a microbe $\boldsymbol{m}_i$ and a drug $\mathbf{d}_j$ have an association relationship, and $A_{i,j}$ is the value '0' otherwise.

For the homogenous similarity graph construction, the original features of microbe nodes include the functional similarity and Gaussian interaction-profile-kernel-based similarity to address the potential

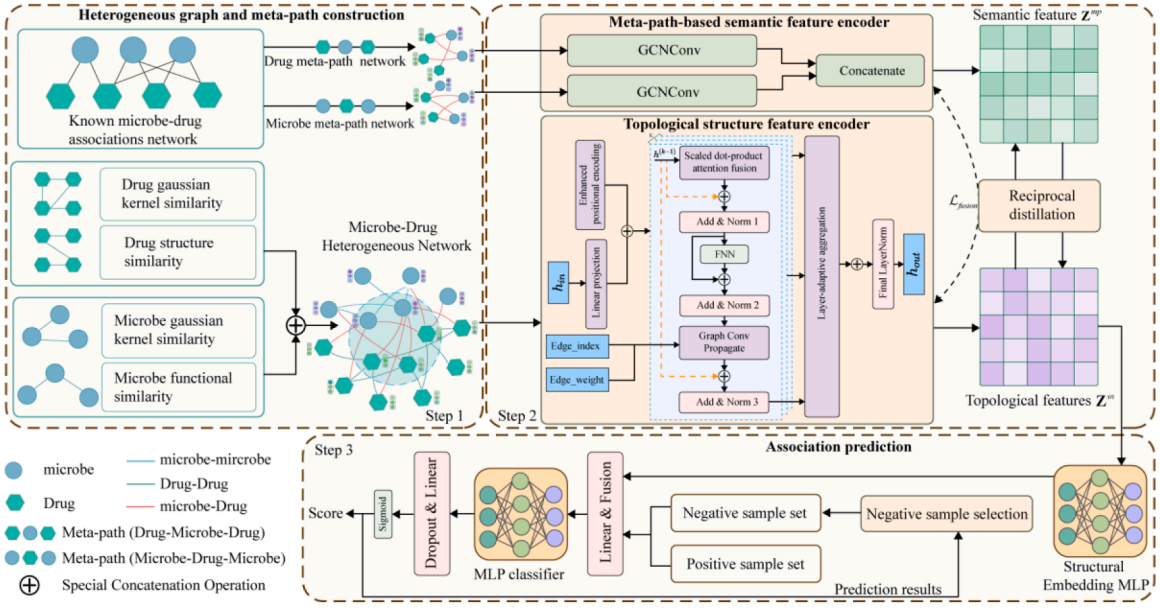


Fig. 1. Illustrative diagram of HGCRL, which includes heterogeneous graph and meta-path construction, topological structure feature encoder, meta-path-based semantic feature encoder, and association prediction. The self-paced negative sampling is used to select informative negative samples to train the classifier.

sparsity in feature-based similarities. We calculated and integrated the similarities of microbe nodes for all the microbe pairs and then constructed the homogenous microbe similarity graph, as Tian et al. did [8]. Then, the original features of drug nodes include the structure similarity and Gaussian interaction-profile-kernel-based similarity, and we calculated and integrated the similarities of drugs for all the drug pairs, as Tian et al. did [8], and then constructed the homogenous graph of drug interaction networks.

2) Heterogeneous graph and meta-path-based subgraph construction

The association networks between microbes and drugs are a heterogeneous biological information network with a complex structure and rich semantics. Meta-paths could comprehensively represent the structure and semantic information of heterogeneous biological information. Consequently, we established a heterogeneous graph based on a homogeneous similarity graph of drugs and microbes, and meanwhile meta-path-based subgraph for microbes and drugs according to the diverse meta-paths. Specifically, through specific sequences of node and edge types, the meta-paths are defined to capture higher-order relationships with distinct semantic information [37]. First, the weighted heterogeneous graph was constructed to integrate the original features S_m of microbes, the original features S_d of drugs, and microbe-drug associations A.

In this work, two meta-paths are designed to establish a meta-path-based subgraph for drugs and microbes. The first meta-path is defined as $M \rightarrow^{R_{MD}} D \rightarrow^{R_{DM}} M$ and abbreviated as MDM, where M is the microbe node and D is the drug node. The second meta-path is defined as $D \rightarrow^{R_{DM}} M \rightarrow^{R_{MD}} D$and abbreviated as DMD. R_{MD}, and R_{DM} represent the relationships of microbe-drug and drug-microbe from the association matrix $\boldsymbol{A}$ and its transpose $\boldsymbol{A}^T$, respectively. For the association matrix $A \in \mathbb{R}^{N_m \times N_d}$ of microbes and drugs, the adjacency matrix of MDM is computed as $A_{MDM} = A \times A^T$, resulting in $A_{MDM} \in \mathbb{Z}_{\geq 0}^{N_m \times N_m}$ and each entry $(A_{MDM})_{ik}$ quantifies the number of shared drugs between microbes $\boldsymbol{m}_i$ and $\boldsymbol{m}_k$. Similarly, the adjacency matrix of DMD is formulated as $A_{DMD} = A^T \times A$ and $A_{DMD} \in \mathbb{Z}_{\geq 0}^{N_d \times N_d}$, where each entry $(A_{DMD})_{jl}$ represents the count of shared microbes between drugs $\boldsymbol{d}_j$ and $\boldsymbol{d}_l$.

3.2. Topological structure feature encoder

Let $\mathbf{X} \in \mathbb{R}^{N \times d_{in}}$ be the initial feature matrix for all N nodes in the graph, where $N = N_m + N_d$ is the total number of microbes and drugs, and d_{in} is the dimension of their initial features. The initial vectors of all the nodes are organized into a matrix $\mathbf{H}_{init}$ based on the association matrix $A \in \mathbb{R}^{N_m \times N_d}$ of microbes and drugs. The initial features of microbes and drugs belong to different spaces, and they are transformed into the same vector space. Then, the node features $\mathbf{H}_{init}$ are transformed based on a nonlinear transformation, and the calculation of node features $\mathbf{H}^{(0)}$ is formulated as $\mathbf{H}^{(0)} = \upsilon(\mathbf{H}_{init}\mathbf{W}_0 + \mathbf{b}_0)$, where υ is the exponential linear unit, and $\mathbf{W}_0 \in \mathbb{R}^{d_{in} \times d}$ is a learnable weight matrix, and $\mathbf{b}_0 \in \mathbb{R}^{1 \times d}$ is a learnable bias vector. The dimension of $\mathbf{H}^{(0)}$ may be consistent with the dimension of position information extracted by the positional information extraction module. Thus, $\mathbf{H}^{(0)}$ is passed into the linear transformation to obtain the same dimension features $\mathbf{H}_p^{(0)}$, and the process is formulated $\mathbf{H}_p^{(0)} = \mathbf{H}^{(0)}\mathbf{W}_p + \mathbf{b}_p$, where W_p is a learnable weight matrix and b_p is a learnable bias vector. The topological structure feature encoder is composed of a layer-adaptive graph transformer (LAGT), and mainly includes positional information extraction, layer-adaptive graph transformer mechanisms, and layer-adaptive aggregation of node information.

1) Positional information extraction

To iteratively refine the discriminative features of drug and microbe nodes, a positional information extraction module is devised based on the static positional encoding and learnable positional encoding. As shown in Fig. 1, we adopted a gated control mechanism to combine different types of node information and obtain the node representation $\mathbf{PE}_f \in \mathbb{R}^{N \times d}$ with positional embedding, and the calculation process is formulated as:

$$\mathbf{PE}_f = \mathbf{G} \odot \mathbf{PE}_s + (1 - \mathbf{G}) \odot \mathbf{PE}_l, \tag{1}$$

where $\odot$ denotes element-wise multiplication. $\mathbf{PE}_s$ is a static positional encoding feature, and $\mathbf{PE}_l$ is a dynamic positional encoding feature. $\mathbf{G}$ is a dynamic gate control mechanism that allows the model to extract the specific positional information of microbes and drugs. The control ratio function of $\mathbf{G}$ is calculated by the formula:

$$\mathbf{G} = \sigma\left([\mathbf{PE}_s \| \mathbf{PE}_l]\mathbf{W}_g + \mathbf{b}_g\right), \tag{2}$$

where $\|$ denotes the concatenation operation along the last channel of representation vectors. $\mathbf{W}_g \in \mathbb{R}^{2d \times d}$ and $\mathbf{b}_g \in \mathbb{R}^{1 \times d}$ are the weight matrix and bias vector for the gate control transformation, respectively. σ is the Sigmoid function that controls the ratio of information passing into the next layer, and its output is a value ranging from [0, 1].

For a node at a given position, its static positional encoding vector $\mathbf{PE}_s \in \mathbb{R}^d$ is generated using the standard sinusoidal functions of varying frequencies. Since heterogeneous graphs of microbes and drugs lack a canonical node ordering, we leverage a node's index within the feature matrix as its unique "position". The value for each dimension of the positional encoding vector is calculated as follows:

$$\mathbf{PE}_{s(pos,i)} = \begin{cases} \sin\left(pos/10000^{2i/d}\right) & \text{ifiiseven} \\ \cos\left(pos/10000^{2i/d}\right) & \text{ifiisodd} \end{cases}, \tag{3}$$

where pos is the position of a microbe or drug node, d is the dimension of the embedding vector, and i is the dimension index ($0 \leq i < d/2$). This formulation ensures that each dimension of the positional encoding corresponds to a sinusoid, assigning a unique and universal structural signature to each node based on its position.

Then, the topological information is combined with the initial features of nodes to obtain the position-aware representations $\mathbf{H}^{(0)}$, and the process is formulated as $\mathbf{H}^{(0)} = \mathbf{H}_p^{(0)} + \mathbf{PE}_f$.

2) Layer-adaptive graph transformer

As shown in the LAGT of Fig. 1, we implemented the layer-adaptive transformer mechanisms to extract complex topological and semantic patterns of microbes and drugs based on position-aware representations $\mathbf{H}^{(0)}$. The LAGT architecture is composed of multiple identical layers stacked iteratively. Each layer is designed to refine the node representations by processing them through a sequence of two main modules: a scaled dot-product attention fusion (SDAF) module, followed by a node information aggregation (NIA) module.

The SDAF module is responsible for capturing feature dependencies using our enhanced attention mechanism. The NIA module then further processes these representations by applying a position-wise feed-forward network and aggregating information from the local graph topology. Let $\mathbf{H}^{(k)}$ denote the representation matrix output by the k-th LAGT layer. The overall transformation within a single layer can be conceptually described as:

$$\mathbf{H}^{(k)} = f_{NIA}\left(f_{SDAF}\left(\mathbf{H}^{(k-1)}\right)\right), \tag{4}$$

where f_{SDAF} and f_{NIA} are the transformation functions of their respective modules. In the following sections, we will detail the specific operations within each of these modules.

The scaled dot-product attention fusion (SDAF) module combines the advantages of edge-specific and global node attention to capture interaction dependencies of drug and microbe nodes. It enhances the standard multi-head attention mechanism by injecting vital topological features through two key components: edge-specific attention and global node attention. Specifically, the input feature of the SDAF module, the node representation from the previous layer $\mathbf{H}^{(k-1)} \in \mathbb{R}^{N \times d}$, is

first projected into query, key, and value matrices for H parallel attention heads. For the h-th head, the projections are: $\mathbf{Q}_h = \mathbf{H}^{(k-1)}\mathbf{W}_h^Q$, $\mathbf{K}_h = \mathbf{H}^{(k-1)}\mathbf{W}_h^K$, $\mathbf{V}_h = \mathbf{H}^{(k-1)}\mathbf{W}_h^V$, where $\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V \in \mathbb{R}^{d\times d_h}$ are learnable, head-specific weight matrices, and $d_h = d/H$.

Then, edge-specific attention captures the pairwise interaction between microbes and drugs. For edge (i,j) and head h, a head-specific scalar weight $w_{ij,h}^{(k)} \in \mathbb{R}$ is computed by a shared multilayer perceptron, and the process is formulated as $w_{ij,h}^{(k)} = \left[f^e\left(\left[\mathbf{h}_i^{(k-1)} \| \mathbf{h}_j^{(k-1)}\right]\right)\right]_h$, where f^e outputs a vector of dimension H, and $(\cdot)_h$ selects the h-th scalar element. This weight then modulates the standard scaled dot-product attention. The attention probability for the head h is formulated as:

$$\alpha_{ij,h}^{(k)} = \mathrm{softmax}_{j\in\mathcal{N}(i)\cup\{i\}}\left(\frac{\left(\mathbf{q}_{i,h}^{(k)}\right)^\top \mathbf{k}_{j,h}^{(k)}}{\sqrt{d_h}} \cdot w_{ij,h}^{(k)}\right) \tag{5}$$

where $\mathbf{q}_{i,h}^{(k)}$ and $\mathbf{k}_{j,h}^{(k)}$ are the query and key vectors for nodes i and j in the head h, respectively. The output $\mathbf{head}_{i,h}^{(k)}$ of a node from the head h is an aggregation of value vectors from neighbor nodes:

$$\mathbf{head}_{i,h}^{(k)} = \sum\nolimits_{j\in\mathcal{N}(i)\cup\{i\}} \alpha_{ij,h}^{(k)} \mathbf{v}_{j,h}^{(k)} \tag{6}$$

Next, the global node attention assesses the overall structural importance of the microbe and drug node. A gate control score $g_i^{(k)}$ is calculated by the formula: $g_i^{(k)} = \sigma\left(f^g\left(\mathbf{h}_i^{(k-1)}\right)\right)$. After the outputs for all attention heads are concatenated into a single d-dimensional vector, this scalar gate $g_i^{(k)}$ is broadcast to match its dimensions and applied via element-wise multiplication. This step allows the model to scale the entire feature representation of a node based on its perceived importance. The final output vector $\mathbf{o}_i^{(k)}$ for the node i is computed by the formula:

$$\mathbf{o}_i^{(k)} = f_{\mathrm{SDAF}}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \left(\| \left(\mathrm{head}_{i,1}^{(k)}, \ldots, \mathrm{head}_{i,\mathrm{H}}^{(k)}\right) \odot g_i^{(k)}\right)\mathbf{W}_o, \tag{7}$$

Where $\|$ is the concatenation operation. $g_i^{(k)}$ is broadcast and applied element-wise to the concatenated attention head vector, which is a scalar multiplier. $\mathbf{W}_o \in \mathbb{R}^{d\times d}$ is the output projection matrix. The final node representations of drugs and microbes are integrated via the residual connection and layer normalization. The calculation function is formulated as $\mathbf{H}_{sdaf}^{(k)} = g_{ln}\left(\mathbf{H}^{(k-1)} + \mathrm{Dropout}\left(\mathbf{O}^{(k)}\right)\right)$, where $\mathbf{O}^{(k)}$ stacks all outputs $\mathbf{o}_i^{(k)}$ of nodes. g_{ln} is the layer normalization function, and solves the stability problem caused by the change in data distribution during training.

3) Layer-adaptive aggregation of node information

The position-wise fully connected network (PFN) is applied independently to each representation of nodes to provide further nonlinear transformation, and the PFN function consists of two nonlinear transformations, and the calculation is formulated as:

$$\mathbf{O}_{pfn}^{(k)} = \vartheta\left(\mathbf{H}_{sdaf}^{(k)}\mathbf{W}_1^{(k)} + \mathbf{b}_1^{(k)}\right)\mathbf{W}_2^{(k)} + \mathbf{b}_2^{(k)}, \tag{8}$$

where ϑ is a rectified linear unit function. $\mathbf{W}_1^{(k)} \in \mathbb{R}^{d\times d_{ffn}}$, $\mathbf{b}_1^{(k)} \in \mathbb{R}^{1\times d_{ffn}}$, $\mathbf{W}_2^{(k)} \in \mathbb{R}^{d_{ffn}\times d}$, and $\mathbf{b}_2^{(k)} \in \mathbb{R}^{1\times d}$ are the layer-specific learnable parameters for the PFN. The transformation function uses different parameters from layer to layer across different positions. The output is processed with residual connection and layer normalization $\mathbf{H}_{pfn}^{(k)} = g_{\ln}\left(\mathbf{H}_{sdaf}^{(k)} + \mathrm{Dropout}\left(\mathbf{O}_{pfn}^{(k)}\right)\right)$.

Then, a graph convolution explicitly integrates local graph topology. First, the refined representations from the FFN are linearly transformed to create features for propagation: $\mathbf{M}^{(k)} = \mathbf{H}_{pfn}^{(k)}\mathbf{W}_{agc}^{(k)}$. These features are then aggregated across the graph structure. The neighborhood aggregation is performed via matrix multiplication with the graph's adjacency matrix $\mathbf{A}$. The aggregated information matrix is thus $\mathbf{AM}^{(k)}$. To ensure unimpeded gradient flow and training stability, this aggregated information is fused with the node representations from the previous layer via a residual connection. The final topological representations $\mathbf{H}^{(k)}$ for the k-th layer are obtained as:

$$\mathbf{H}^{(k)} = g_{ln}\left(\mathbf{H}^{(k-1)} + \mathrm{Dropout}\left(\mathbf{AM}^{(k)}\right)\right) \tag{9}$$

where g_{ln} is the layer normalization function.

Next, the output of each internal layer propagates as input to the subsequent layer for iterative transformation, and simultaneously generates a repository $C = \left[\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(K)}\right]$ from varying receptive fields. By capturing information at multiple scales, the model preserves distinct node representations across layers. Shallower layers retain local details, while deeper layers integrate broader contextual information. Such multi-scale aggregation mitigates the loss of node-specific characteristics during repeated propagation, ensuring that features remain distinguishable even after multiple iterations [38].

After all K layers have been executed, the layer-adaptive aggregation mechanism synthesizes the multi-scale features $C = \left[\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(K)}\right]$. The fusion operation dynamically assigns the importance of each layer by the weights $\beta_k = \mathrm{softmax}\left(\alpha_f\right)_k$ from a learnable parameter vector $\alpha_f \in \mathbb{R}^K$:

$$\mathbf{Z}^{sn} = g_{ln}\left(\sum_{k=1}^{K} \beta_k \cdot \mathbf{H}^{(k)}\right), \tag{10}$$

where g_{ln} is the layer normalization function. Unlike graph attention networks (GAT) or GraphSAGE [39,40], which typically rely only on the last-layer output $\mathbf{H}^{(K)}$, this strategy leverages information across multiple depths. Discarding earlier representations makes standard models prone to feature convergence, as repeated neighbor averaging blurs node distinctions [41]. In contrast, our aggregation preserves and adaptively fuses multi-scale features, maintaining representational diversity and mitigating the exponential decay of feature variance that commonly arises in deep graph convolutions [42].

3.3. Meta-path-based semantic feature encoder

The semantic feature extraction module adopts two distinct meta-paths to learn representations for microbes and drugs from their corresponding homogeneous subgraphs. This model allows the model to capture specific high-order structural information for each node type. As the feature learning process is analogous for both microbes and drugs, we will use microbes as a detailed example to illustrate the methodology. The process begins by taking the initial feature matrix of microbes $\mathbf{H}_m \in \mathbb{R}^{N_m\times d}$ as input. The graph structure for the operation is derived from the MDM meta-path, resulting in a microbe-microbe adjacency matrix $\mathbf{A}_{MDM}$. A graph convolutional layer [43], enhanced with a residual connection, is then applied to learn the semantic representations of the microbe nodes. The entire transformation to generate the final semantic embedding matrix for microbes $\mathbf{Z}_m^{sn}$ is defined as:

$$\mathbf{Z}_m^{sn} = \delta\left(\mathrm{Dropout}\left(\left(\widetilde{\mathbf{D}}_{MDM}^{-\frac{1}{2}}\widetilde{\mathbf{A}}_{MDM}\widetilde{\mathbf{D}}_{MDM}^{-\frac{1}{2}}\mathbf{H}_m\mathbf{W}_m\right) + \mathbf{H}_m + \mathbf{b}_m\right)\right), \tag{11}$$

where $\widetilde{\mathbf{A}}_{MDM} = \mathbf{A}_{MDM} + \mathbf{I}$ is the adjacency matrix with self-loops added, and $\widetilde{\mathbf{D}}_{MDM}$ is its corresponding diagonal degree matrix. $\mathbf{W}_m$ and $\mathbf{b}_m$ are

the learnable weight matrix and bias vector for the microbe-specific transformation, respectively. δ represents a parametric rectified linear unit function. The operation involves a sequence of steps: graph convolution, addition of a residual connection and the bias term, followed by a Dropout layer for regularization. To effectively regularize and prevent the co-adaptation of neurons, during training, the output elements of graph convolutions are randomly zeroed with a probability and the zeroed elements are chosen independently and are sampled from a Bernoulli distribution.

$$\begin{cases} \mathbf{Z}_{\text{enc}}^{sn} = f_{enc}^{sn}(\mathbf{Z}^{sn}) = \vartheta\left(g_{ln}\left(\mathbf{Z}^{sn}\mathbf{W}_{enc,1}^{sn} + \mathbf{b}_{enc,1}^{sn}\right)\right)\mathbf{W}_{enc,2}^{sn} + \mathbf{b}_{enc,2}^{sn} \\ \mathbf{Z}_{enc}^{mp} = f_{enc}^{mp}(\mathbf{Z}^{mp}) = \vartheta\left(g_{ln}\left(\mathbf{Z}^{mp}\mathbf{W}_{enc,1}^{mp} + \mathbf{b}_{enc,1}^{mp}\right)\right)\mathbf{W}_{enc,2}^{mp} + \mathbf{b}_{enc,2}^{mp} \\ \mathbf{Z}_{proj} = f_{proj}\left(\mathbf{Z}_{enc}^{mp}\right) = \vartheta\left(g_{ln}\left(\vartheta\left(g_{ln}\left(\mathbf{Z}_{proj}\mathbf{W}_{proj,1}^{mp} + \mathbf{b}_{proj,1}^{mp}\right)\right)\mathbf{W}_{proj,2}^{mp} + \mathbf{b}_{proj,2}^{mp}\right)\mathbf{W}_{enc,3}^{mp} + \mathbf{b}_{enc,3}^{mp}\right) \\ \mathbf{Z}_{pred} = f_{pred}\left(\mathbf{Z}_{proj}\right) = \vartheta\left(g_{ln}\left(\mathbf{Z}_{proj}\mathbf{W}_{pred,1}^{mp} + \mathbf{b}_{pred,1}^{mp}\right)\right)\mathbf{W}_{pred,2}^{mp} + \mathbf{b}_{pred,2}^{mp} \end{cases}, \tag{12}$$

Similarly, the semantic embeddings for drug nodes $\mathbf{Z}_d^{mp}$ are extracted through an analogous process based on their unique DMD meta-path subgraph. To form a unified representation, the resulting semantic feature matrices for all drugs and microbes are $\mathbf{Z}^{mp} = \| \left(\mathbf{Z}_d^{mp}, \mathbf{Z}_m^{mp}\right)$, which are extracted by the meta-path-based semantic feature encoder.

3.4. Reciprocal distillation-based optimization

(1) Reciprocal distillation module of microbes and drugs

In this work, inspired by SimSiam [44], we devised an asymmetric reciprocal distillation module (RDM) to enhance heterogeneous graph collaborative representation learning. By adapting SimSiam, which relies on positive pairs alone to learn invariant representations, the RDM promotes effective knowledge transfer across modalities and strengthens robustness against data heterogeneity and imbalances [45]. In this work, the RDM module leads to embeddings of microbes and drugs that not only remain stable, but also maintain consistency between modalities by aligning their feature spaces without reliance on negative samples.

The RDM module includes a hybrid weight-sharing strategy and an asymmetric distillation pipeline to combine structural features $\mathbf{Z}^{sn} \in \mathbb{R}^{N\times d}$ with topological features$\mathbf{Z}^{mp} \in \mathbb{R}^{N\times d}$ for microbe and drug nodes. The RDM module employs a specific topological encoder f_{enc}^{sn} and semantic encoder f_{enc}^{mp} to capture the intrinsic characteristics of microbes and drugs, as shown in Fig. 2. Each encoder is a two-layer MLP with input dimension d and output dimension h based on linear layers, layer normalization, and ReLU activation. Then, the information transfer-aided matching algorithm leverages both node features of global and local perspectives to enhance the alignment between drug and microbe representations, and the feature outputs of the special encoders are mapped into a common latent space by a shared projectorf_{proj}. The projector is a three-layer MLP with input dimension h and output dimension p. The feature calculation of a specific topological encoder f_{enc}^{sn}, a semantic encoder f_{enc}^{mp} and a shared projector f_{proj} are formulated as:

Besides, after extracting topological and semantic features, the cross-level alignment is needed for efficient information transfer. The RDM module consists of two distinct optimization pathways: 1) The online pathway is actively trained via backpropagation and includes the modality-specific encoders, the shared projector, and a shared predictor f_{pred}, which transforms the online projection of one view to match the target projection of the opposing view. 2) The target pathway provides stable regression targets, and mirrors the online pathway's encoders and projector but lacks a predictor. This dual-pathway architecture, particularly the asymmetry introduced by the predictor and the stop-gradient mechanism, is crucial for stabilizing the training dynamics without contrastive pairs [45]. It fosters a meaningful cross-view alignment where one view can learn from a consistent target provided by the other. Besides, the parameters θ' of RDM are not updated by gradients but are instead maintained as an exponential moving average (EMA) of the online parameters θ governed through the rule:$\theta' \leftarrow \tau\theta' + (1-\tau)\theta$, where τ is the momentum coefficient [46]. This dual-pathway architecture uses asymmetry and momentum-based updates to foster meaningful cross-view alignments, while the stop-gradient mechanism helps maintain distinct yet complementary representations, thereby improving convergence and preserving modality-specific details in heterogeneous contexts [47]. This gradual alignment process enables shared-space learning while retaining modality-specific nuances, which complements the variance–covariance regularization introduced later and results in richer cross-modal representations.

Compared to self-distillation with no labels(DINO), the RDM adapts this for graphs by emphasizing reciprocal exchanges across heteroge-

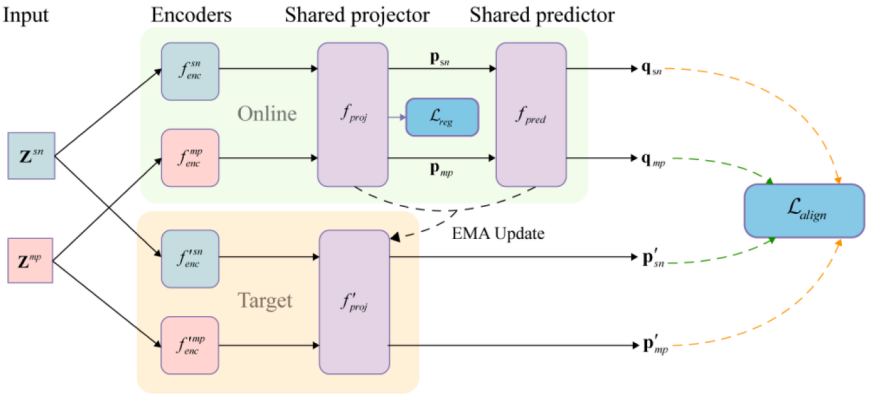


Fig. 2. The flowchart of the reciprocal distillation module.

neous features, enabling finer handling of node interactions without large-scale augmentations [48]. Compared to bootstrap your own latent (BYOL), our module incorporates explicit regularization to further decorrelate features, enhancing performance in sparse or imbalanced networks [49]. Moreover, the RDM module is optimized via a compound loss function $\mathscr{L}_{fusion}$, which linearly combines the special objectives of a cross-level alignment loss and a latent space regularization loss. The cross-level alignment loss $\mathscr{L}_{align}$ is the primary objective driving the reciprocal distillation process. $\mathscr{L}_{align}$ is a symmetric negative cosine similarity loss that compels the topological-level pathway prediction:

$$\mathscr{L}_{align} = \frac{1}{2}\left[\mathscr{D}\left(\mathbf{q}_{sn}, \mathbf{p}'_{mp}\right) + \mathscr{D}\left(\mathbf{q}_{mp}, \mathbf{p}'_{sn}\right)\right], \tag{13}$$

where $\mathbf{q}_{view} \in \mathbb{R}^{N\times p} = f_{pred}\left(f_{proj}\left(f_{enc}^{view}\left(\mathbf{Z}^{view}\right)\right)\right)$ is the topological-level pathway prediction, $p'_{view} = f'_{proj}\left(f_{enc}^{view}\left(\mathbf{Z}^{view}\right)\right)$ is the semantic-level projection, and $\mathscr{D}$ denotes the negative cosine similarity.

(2) Loss function

Inspired by variance-invariance-covariance regularization (VICReg) [50], we introduced a latent space regularization loss $\mathscr{L}_{reg}$ that operates directly on the topological structure projections ($\mathbf{p}_{sn}, \mathbf{p}_{mp} \in \mathbb{R}^{N\times p}$). This regularization component directly enforces two critical statistical properties on the projected embeddings, such as variance preservation and covariance decorrelation [50]. The former prevents the embeddings from collapsing to a trivial solution, while the latter reduces feature redundancy. This is particularly beneficial for maintaining informative and disentangled representations in a shared multi-modal space. To ensure the shared latent space and avoid the dimensional collapse, the invariance term of the original VICReg is omitted as this objective is already fulfilled by $\mathscr{L}_{align}$.$\mathscr{L}_{reg}$ is composed of a variance term $\mathscr{L}_{var}$ and a covariance term $\mathscr{L}_{cov}$. The calculation of $\mathscr{L}_{reg}$ is formulated as:

$$\mathscr{L}_{reg} = \sum_{view\in\{sn,mp\}} \left(\mu_{var}\mathscr{L}_{var}(\mathbf{p}_{view}) + \nu_{cov}\mathscr{L}_{cov}(\mathbf{p}_{view})\right), \tag{14}$$

where $\mathscr{L}_{var}$encourages the variance of embeddings along each feature dimension to be close to a target value, thus preserving information. $\mathscr{L}_{cov}$ decorrelates different feature dimensions, reducing redundancy. Thus, the feature fusion loss of microbe and drugs is $\mathscr{L}_{fusion} = \gamma_{align}\mathscr{L}_{align} + \gamma_{reg}\mathscr{L}_{reg}$, where hyperparameters γ_{align} and γ_{reg} balance the contributions of alignment and regularization. The fusion loss integrates cross-modal alignment and latent space regularization objectives, encouraging effective information exchange between topological-level and semantic-level learning pathway. The complete training procedure for the RDM is detailed in Algorithm 1.

Algorithm 1
The Reciprocal Distillation Module (RDM).

Inputs:
Embeddings: $\mathbf{Z}^{sn}, \mathbf{Z}^{mp}$
Online networks: $\{f_{enc}^{sn}, f_{enc}^{mp}, f_{proj}, f_{pred}\}$ with parameters θ
Target networks: $\left\{f_{enc}^{sn}, f_{enc}^{mp}, f'_{proj}\right\}$ with parameters θ'
Hyperparameters: EMA decay τ, loss weights $\gamma_{align}, \gamma_{reg}, \mu_{var}, \nu_{cov}$
Train Step:
1: Initialize online parameters θ; copy to target parameters $\theta' \leftarrow \theta$
2: **for** each training iteration **do**
3: // forward pass for both views $v \in \{sn, mp\}$
4: // Online forward pass
5: $\mathbf{p}_v \leftarrow f_{proj}\left(f_{enc}^{v}\left(\mathbf{Z}^{v}\right)\right)$
6: $\mathbf{q}_v \leftarrow f_{pred}\left(\mathbf{p}_v\right)$
7: // Target pass with no gradients (sg = stop-gradient)
8: $P'_v \leftarrow \mathrm{sg}\left(f'_{proj}\left(f_{enc}^{v}\left(Z^{v}\right)\right)\right)$
9: // compute loss
10: $\mathscr{L}_{align} \leftarrow 0.5 * \left[\mathscr{D}\left(\mathbf{q}_{sn}, \mathbf{p}'_{mp}\right) + \mathscr{D}\left(\mathbf{q}_{mp}, \mathbf{p}'_{sn}\right)\right]$
11: $\mathscr{L}_{reg} \leftarrow \sum_v \left(\mu_{var}\mathscr{L}_{var}(\mathbf{p}_v) + \nu_{cov}\mathscr{L}_{cov}(\mathbf{p}_v)\right)$
12: $\mathscr{L}_{fusion} \leftarrow \gamma_{align}\mathscr{L}_{align} + \gamma_{reg}\mathscr{L}_{reg}$
13: // update parameters
14: update θ via backpropagation on $\mathscr{L}_{fusion}$
15: update $\theta' \leftarrow \tau\theta' + (1-\tau)\theta$
16: **end for**

To effectively address the inherent sample imbalance, as they [51] did, we designed an efficient loss function, which is denoted as a focal and distillation (FD) loss function. The novel optimization loss combined the focal loss and feature fusion loss of microbes and drugs to capture the structural and semantic information of nodes. The focal loss assigns higher attention to hard-to-classify samples while down-weighting the loss contribution from easily classified ones. The calculation of the FD loss is defined as:

$$\mathscr{L}_{\mathrm{fd}} = -\alpha_t(1-p_t)^{\gamma}\log(p_t) + \lambda_{ctr}\mathscr{L}_{fusion}, \tag{15}$$

where $\mathscr{L}_{focal}$ is the primary task loss, and $\mathscr{L}_{fusion}$ is the fusion loss from the RDM module. λ_{ctr} balances the contribution of the fusion objective. p is the estimated probability of models for the positive label. The term $\gamma \geq 0$ is a tunable focusing parameter. α_t is a balancing parameter, which is a hyperparameter to balance the importance of positive/negative samples. The joint optimization enables the model to learn discriminative and robust features through reciprocal knowledge distillation, while simultaneously enhancing performance on imbalanced classification tasks.

3.5. Association prediction

To derive final embeddings of nodes, the enhanced structural embeddings are integrated by a multi-layer perceptron (MLP), and layer normalization (LN). The transformation is formulated as $\mathbf{Z}^{f} = g_{ln}^{f}\left(f_{mlp}^{f}(\mathbf{Z}^{sn})\right)$, where f_{mlp}^{f} is a calculation process of a multi-layer perceptron, and g_{ln}^{f} is a calculation process of layer normalization in the association predictor. Then, the ultimate embeddings for microbes and drugs can be represented as $\mathrm{Z}_{m_i}^{f}$ and$\mathrm{Z}_{\mathrm{d}_j}^{f}$ based on MLP and LN. The representation features of microbes and drugs are combined by the element-wise multiplication to form the complete fusion representation $\mathbf{z}_{pair}$:

$$\mathbf{z}_{pair} = \left(f_{m}^{linear}\left(\mathrm{Z}_{m_i}^{f}\right)\right) \odot \left(f_{d}^{linear}\left(\mathrm{Z}_{\mathrm{d}_j}^{f}\right)\right), \tag{16}$$

where f_{m}^{linear} and f_{d}^{linear} are one-layer linear transformation to obtain representations in the same dimension space for drug and microbes.

Next, $\mathbf{z}_{pair}$ is subsequently fed into a prediction layer f_{mlp}. The prediction layer consists of a linear transformation layer with layer normalization, a rectified linear unit function, and dropout optimization. The low-dimensional informative embedding of drugs and microbes is then passed to an output layer, and the output layer maps the features to a possibility p_{ij} of whether there exists association information between a drug and a microbe as $p_{ij} = \sigma\left(f^{\mathrm{linear}}\left(f_{mlp}\left(\mathbf{z}_{pair}\right)\right)\right)$, where f_{mlp} represents linear transformation of two layers with rectified linear unit function and f^{linear} is the linear transformation layer. σ is the Sigmoid function that controls the ratio of information passing into the next layer, and its output a value ranging from [0, 1].

4. Results and discussions

4.1. Experiment settings and evaluation metrics

This section introduces the experimental settings and evaluation metrics, respectively. The implementation of all models was developed by Python 3.10.8, PyTorch 2.1.2, and PyTorch Geometric 2.5.3. All experiments were executed on a server equipped with an Intel Xeon Silver 4214R processor and an NVIDIA RTX 3080Ti GPU. The parameters of models were initialized by the Glorot scheme and optimized based on the Adam algorithm. Moreover, we used a cosine annealing

schedule to dynamically adjust the learning rate, which was initially set to 0.0006 and decayed over 200 epochs to a lower bound of 0.00001. The LAGT module incorporated a 4-head self-attention mechanism. The embedding dimensions for both microbes and drugs, as well as the dimensions of all hidden layers, were uniformly set to 128. All models were trained for a total of 1000 epochs. For the computational cost of our model, we recorded the total training time and resource consumption for the complete 5-fold cross-validation procedure. For the MDAD dataset, the training was completed in 803 s, with a peak GPU memory consumption of approximately 4.17 GB and system memory usage around 1.66 GB. For the aBiofilm dataset, the total training time was 1387s, with peak GPU and system memory usages of 8.76 GB and 1.7 GB, respectively. Finally, the experiments on the DrugVirus dataset took 424 s, consuming around 0.21 GB of GPU memory and 1.5 GB of system memory.

Then, we selected the average area under the receiver operating characteristic curve (AUC), and area under the precision-recall curve (AUPRC) metrics to evaluate models, as discussed in a previous study [52,53]. Moreover, the model performance was also assessed by Accuracy (ACC), F1-score, and Matthews Correlation Coefficient (MCC). To assess the statistical significance of performance differences, we performed paired t-tests across the 5-fold CV results and reported p-values. We calculated and reported 95% confidence intervals (CI) for the key performance metrics, providing a range that likely contains the true population parameter with 95% confidence, thereby offering a measure of precision and uncertainty in our estimates.

All the known microbe-drug association pairs are treated as positive samples and form the positive sample set, whereas all the remaining unknown microbe-drug association pairs are regarded as candidate negative samples and form the candidate negative sample set. For negative sample selection, we adopt a sample selection strategy similar to that used in [53], in which negative samples are drawn from the candidate set at different ratios (1:1, 1:5, 1:10) relative to the positive samples.

4.2. Baseline methods and performance comparison

1) Baseline methods

The section introduces the baseline models as follows:

SCSMDA [8] employs a structure-enhanced contrastive learning strategy to learn representations from both similarity and meta-path-induced biological graphs.

MDASAE [28] is a deep microbe-drug association prediction model that combines stacked autoencoders with a multi-head attention mechanism. It integrates disease-related information and employs a random walk with restart algorithm to predict associations.

GARFMDA [54] integrates a graph attention network with a bilayer random forest to infer potential microbe-drug associations, which leverage drug and microbe features from multi-source similarity networks

OGNNMDA [55] introduces an ordered message-passing graph neural network to preserve layer-distinct information of neighbor nodes.

GCNATMDA [27] is a hybrid model that integrates graph convolutions to extract latent entity features and graph attention to learn complex interaction relationships for association prediction.

NRGCNMDA [2] is an integrated association prediction model that combines Node2vec with a residual-enhanced graph convolution for learning high-order features.

CLMT [13] combines a transformer-based graph attention with multi-view graph contrastive learning to capture high-order graph dependency patterns under conditions of data sparsity.

KNDM [14] predicts drug-microbe associations by combining a knowledge graph transformer with node category-sensitive contrastive learning to encode contextual relationships from multiple meta-paths.

GTFKAN [15] is a novel microbe-drug association model by using a graph transformer for node encoding and a feature kernel attention mechanism.

2) Model performance

In this section, we compared the HGCRL model with several remarkable methods on MDAD, aBiofilm, and DrugVirus datasets based on varying class imbalance ratios. The results of our model are reported by the mean and standard deviation, and Fig. 3 shows the corresponding ROC and PR curves during training. The HGCRL model obtains exceptional and highly stable performance across all the benchmark datasets. Specifically, the HGCRL model achieved average AUC values of 0.9943 $\pm$ 0.0010 (95% CI: 0.9929-0.9956), and AUPRC scores of 0.9946 $\pm$ 0.0009 (95% CI: 0.9933-0.9959) on the MDAD dataset, achieved average AUC scores of 0.9956 $\pm$ 0.0011 (95% CI: 0.9945-0.9968), and AUPRC scores of 0.9959 $\pm$ 0.0009 (95% CI: 0.9949-0.9969) on the aBiofilm dataset, and achieved average AUC values of 0.9854 $\pm$ 0.0026 (95% CI: 0.9818-0.9890), and AUPRC scores of 0.9862 $\pm$ 0.0033 (95% CI: 0.9815-0.9909) on the DrugVirus dataset. The consistently high performance with minimal standard deviations underscores the capability of HGCRL to accurately and reliably predict microbe-drug associations, and also validates the effectiveness and generalizability of our model across diverse datasets.

3) Performance comparison

Real-world microbe-drug association data are characterized by a profound class imbalance. The inherent sparsity can significantly bias predictive models and impair the performance of models. To evaluate models comprehensively, we conducted extensive experiments of microbe-drug association prediction based on three different ratios between positive and negative samples, and the ratios of positive and negative samples are '1:1' (balanced datasets), '1:5' (imbalanced datasets), and '1:10' (highly imbalanced datasets), respectively. The performance of the HGCRL model is achieved based on AUC and AUPRC with the mean and standard deviation values. Table 1 reports the comprehensive performance of HGCRL and baseline models on the MDAD, aBiofilm, and DrugVirus datasets.

In Table 1, the performance of HGCRL is consistently superior to the baseline models on positive and negative samples with three ratios, and the remarkable results indicate the effectiveness of our model. For the performance of HGCRL on positive and negative samples with a balanced ratio (1:1), HGCRL achieves the highest AUC score of 0.9943 and an AUPRC score of 0.9946, and obtains an average performance improvement by an AUC score of 2.81% and an AUPRC of 4.97% compared to baseline models on the MDAD dataset. HGCRL achieves the highest AUC score of 0.9963 and AUPRC score of 0.9965, and obtains the average performance improvement by an AUC score of 2.79% and an AUPRC score of 5.16% compared to baseline models on the aBiofilm dataset. HGCRL achieves the highest AUC score of 0.9854 and an AUPRC score of 0.9862, and obtains the average performance improvement by an AUC score of 10.69% and an AUPRC score of 12.57% compared to baseline models on the DrugVirus dataset. With the positive and negative samples of a '1:5' ratio, HGCRL obtains the average performance improvement by an AUC score of 5.67% and an AUPRC score of 25.66% at least compared to baseline models on the MDAD, aBiofilm, and DrugVirus dataset. With the positive and negative samples of a '1:10' ratio, HGCRL obtains the average performance improvement by an AUC score of 3.94% and an AUPRC score of 26.27% at least compared to baseline models on the MDAD, aBiofilm, and DrugVirus datasets.

Through the above comparison, we have the observation for microbe-drug association prediction: ①Compared with SCSMDA [8], NRGCNMDA [2], and CLMT [13], the HGCRL model obtains a remarkable performance improvement, and structural and semantic features enhance the feature capability of microbe and drug nodes, which are extracted by the similarity graphs and meta-path-induced graphs of microbes and drugs.② Compared with MDASAE [28] and GARFMDA [54], the HGCRL model could extract topological representation, semantic representations, and higher-order associations through layer-adaptive graph transformer and pathway-aware graph convolutions. The layer-adaptive graph transformer (LAGT) of HGCRL proves

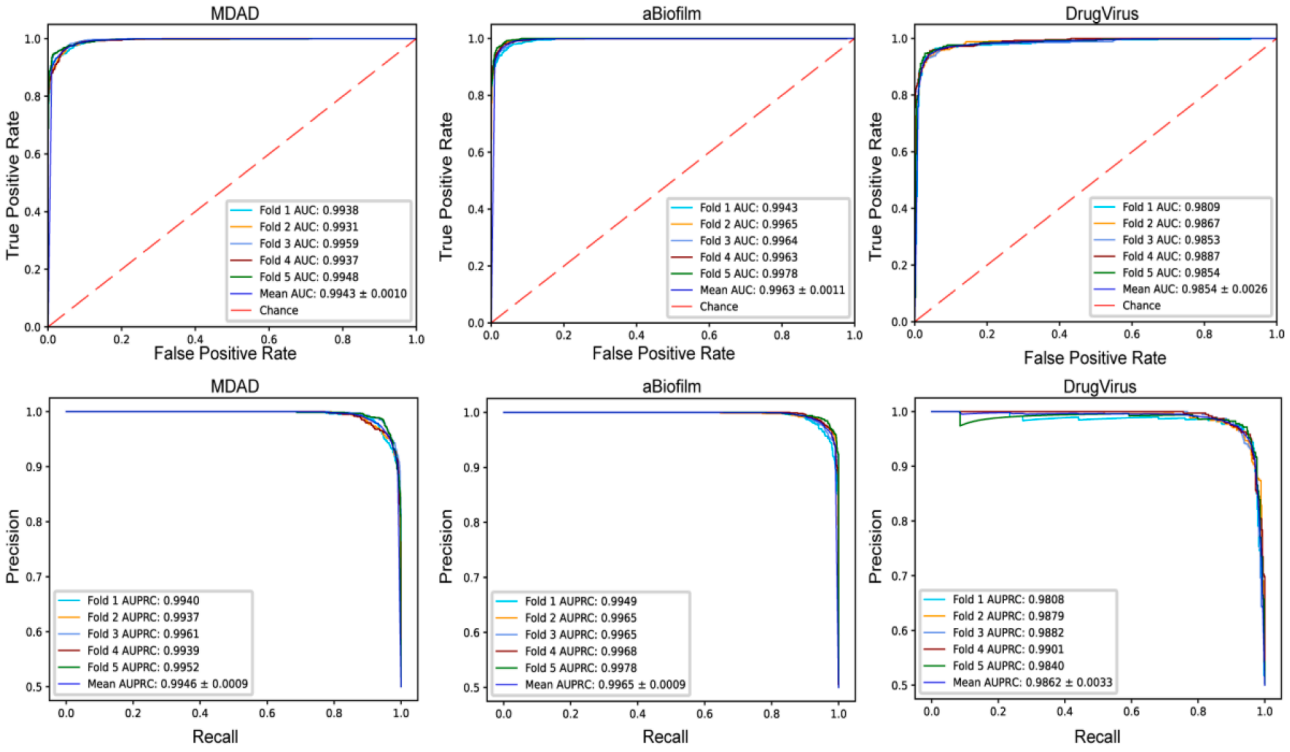


Fig. 3. The AUC and AUPR scores of HGCRL on the MDAD, aBiofilm, and DrugVirus datasets.

Table 1
Performance comparison of HGCRL and baselines. Note: The best results are marked in bold.

Ratios	Models	MDAD		aBiofilm		DrugVirus	
		AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
1:1	SCSMDA	0.9598±0.0014	0.9469±0.0021	0.9651±0.0025	0.9434±0.0035	0.8821±0.0056	0.8617±0.0094
	MDASAE	0.9633±0.0018	0.9647±0.0025	0.9670±0.0022	0.9681±0.0027	0.8558±0.0061	0.8383±0.0090
	GARFMDA	0.9711±0.0015	0.9706±0.0022	0.9755±0.0025	0.9771±0.0030	0.8655±0.0054	0.8672±0.0060
	OGNNMDA	0.9605±0.0020	0.9624±0.0021	0.9673±0.0012	0.9669±0.0009	0.8571±0.0074	0.8612±0.0075
	GCNATMDA	0.9632±0.0043	0.9390±0.0051	0.9663±0.0035	0.9382±0.0045	0.8556±0.0102	0.8614±0.0116
	NRGCNMDA	0.9494±0.0032	0.9325±0.0041	0.9422±0.0051	0.9259±0.0067	0.9278±0.0025	0.8993±0.0036
	CLMT	0.9717±0.0014	0.9710±0.0023	0.9739±0.0023	0.9712±0.0011	0.9706±0.0013	0.9691±0.0013
	KNDM	0.9723±0.0023	0.8693±0.0052	0.9731±0.0020	0.8672±0.0049	0.9037±0.0081	0.8553±0.0136
	GTFKAN	0.9936±0.0034	0.9807±0.0028	0.9960±0.0034	0.9816±0.0028	0.9069±0.0078	0.8833±0.0072
	HGCRL	**0.9943±0.0010**	**0.9946±0.0009**	**0.9963±0.0011**	**0.9965±0.0009**	**0.9854±0.0026**	**0.9862±0.0033**
1:5	SCSMDA	0.9573±0.0020	0.7825±0.0060	0.9621±0.0028	0.7881±0.0060	0.8361±0.0075	0.5463±0.0110
	MDASAE	0.9606±0.0024	0.7872±0.0068	0.9635±0.0025	0.7916±0.0055	0.8695±0.0070	0.5724±0.0110
	GARFMDA	0.9697±0.0021	0.8484±0.0040	0.9747±0.0027	0.8557±0.0039	0.8586±0.0063	0.6847±0.0080
	OGNNMDA	0.9575±0.0022	0.7843±0.0061	0.9642±0.0018	0.7882±0.0062	0.8312±0.0082	0.5528±0.0090
	GCNATMDA	0.9552±0.0049	0.7851±0.0070	0.9632±0.0038	0.7959±0.0056	0.8421±0.0115	0.5757±0.0140
	NRGCNMDA	0.9401±0.0038	0.7562±0.0066	0.9359±0.0055	0.7525±0.0081	0.9152±0.0073	0.6815±0.0092
	CLMT	0.9706±0.0017	0.8476±0.0026	0.9721±0.0029	0.8413±0.0036	0.9236±0.0049	0.7017±0.0081
	KNDM	0.9712±0.0020	0.8411±0.0042	0.9734±0.0024	0.8438±0.0045	0.8936±0.0074	0.7035±0.0093
	GTFKAN	0.9785±0.0020	0.9517±0.0031	0.9749±0.0021	0.9522±0.0026	0.8993±0.0069	0.8173±0.0082
	HGCRL	**0.9916±0.0012**	**0.9729±0.0037**	**0.9933±0.0008**	**0.9768±0.0021**	**0.9693±0.0060**	**0.8842±0.0124**
1:10	SCSMDA	0.9522±0.0022	0.6231±0.0078	0.9575±0.0030	0.6380±0.0070	0.8325±0.0084	0.3974±0.0138
	MDASAE	0.9578±0.0027	0.6820±0.0071	0.9590±0.0029	0.6879±0.0063	0.8673±0.0077	0.4081±0.0120
	GARFMDA	0.9650±0.0023	0.7599±0.0050	0.9702±0.0032	0.7683±0.0047	0.8521±0.0072	0.4199±0.0095
	OGNNMDA	0.9535±0.0026	0.6230±0.0073	0.9616±0.0020	0.6385±0.0071	0.8293±0.0088	0.3991±0.0104
	GCNATMDA	0.9501±0.0059	0.6430±0.0075	0.9602±0.0039	0.6417±0.0067	0.8383±0.0124	0.3990±0.0155
	NRGCNMDA	0.9303±0.0040	0.6226±0.0074	0.9295±0.0074	0.6151±0.0091	0.8544±0.0082	0.4172±0.0101
	CLMT	0.9658±0.0020	0.7375±0.0029	0.9526±0.0036	0.7686±0.0048	0.8795±0.0055	0.4271±0.0097
	KNDM	0.9664±0.0022	0.7392±0.0041	0.9541±0.0033	0.7703±0.0090	0.8762±0.0098	0.4289±0.0105
	GTFKAN	0.9727±0.0020	0.8438±0.0038	0.9596±0.0029	0.8509±0.0071	0.8812±0.0073	0.4304±0.0100
	HGCRL	**0.9904±0.0006**	**0.9536±0.0015**	**0.9822±0.0029**	**0.9091±0.0086**	**0.9041±0.0130**	**0.4584±0.0161**

more effective for drug-microbe association prediction. The integrated edge-specific weights and node-level attention gates allow LAGT to dynamically capture more nuanced topological dependencies.③ Compared with OGNNMDA [55], GCNATMDA [27], the improvement of the HGCRL model can be attributed to our multi-objective optimization strategy. HGCRL does not solely rely on a prediction loss. The addition of the fusion loss acts as a powerful regularization signal, forcing the model to learn better-structured embeddings, while the focal

Loss directly tackles the class imbalance that can impair standard GNN training. ④ Compared with KNDM [14] and GTFKAN [15], HGCRL demonstrates notable improvements. While KNDM leverages knowledge graph transformers with contrastive learning, it shows inconsistent performance across different evaluation metrics. This suggests that KNDM may struggle with the precise ranking of positive associations. For GTFKAN, the sophisticated graph transformer and feature kernel attention mechanism achieve strong performance, but still fall short of HGCRL's results.

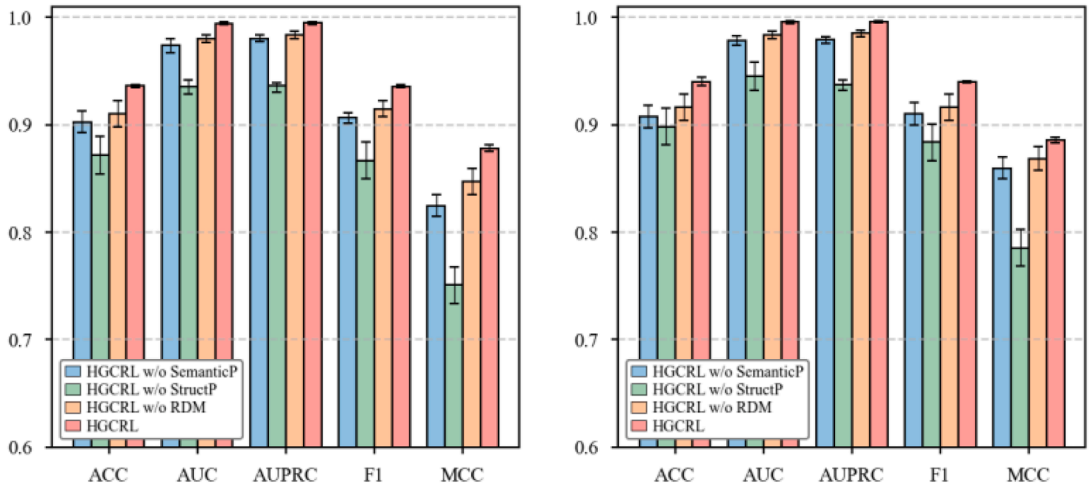
4.3. Ablation analysis

1) Effect of each module

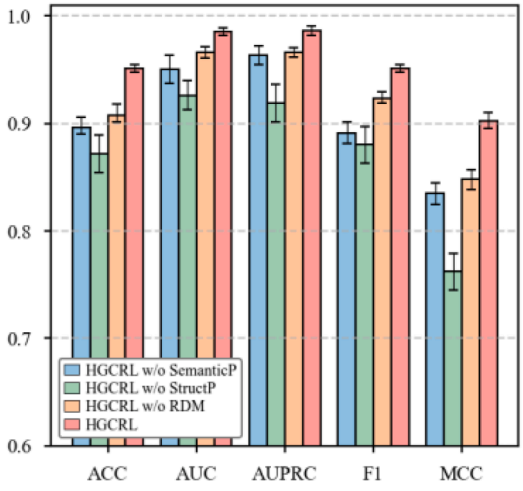
To assess the contribution of each component of HGCRL, we designed the variant models of HGCRL in a comprehensive ablation study as follows:

- HGCRL w/o SemanticP: The variant model removes the semantic pathway (the GCNConv module). Consequently, the reciprocal distillation module (RDM) is also omitted, and the microbe-association predictions are derived solely from the structural representations of the LAGT pathway.
- HGCRL w/o StructP: The variant model removes the primary structural pathway (the LAGT module). The RDM is likewise omitted, and the microbe-association predictions rely exclusively on the semantic representations from the GCNConv pathway.
- HGCRL w/o RDM: The variant model retains both the structural and semantic pathways but deactivates the reciprocal distillation objective within RDM. The microbe-association predictions are conducted based on the unfused representations from the primary LAGT pathway.

Fig. 4 shows the performance of variant models and HGCRL. The HGCRL model consistently and significantly outperformed all variant models across all datasets and metrics. The HGCRL model consistently and significantly outperformed all variant models across all datasets and metrics. The results of the models demonstrate that each component of HGCRL is vital to the efficacy of HGCRL for the microbe-drug association prediction. Specifically, removing the semantic pathway of HGCRL, the average performance of HGCRL dropped by an AUC score of 2.5% and an AUPRC score of 1.8% across the MDAD, aBiofilm, and DrugVirus datasets. Then, the impact of the structural pathway was more pronounced on the performance of HGCRL; removing the structural pathway, the performance of HGCRL dropped by an AUC score of 5.9%



(a) Performance of models on MDAD; (b) Performance of models on aBiofilm



(c) Performance of models on DrugVirus

Fig. 4. The impact of different components on the performance of the HGCRL model.

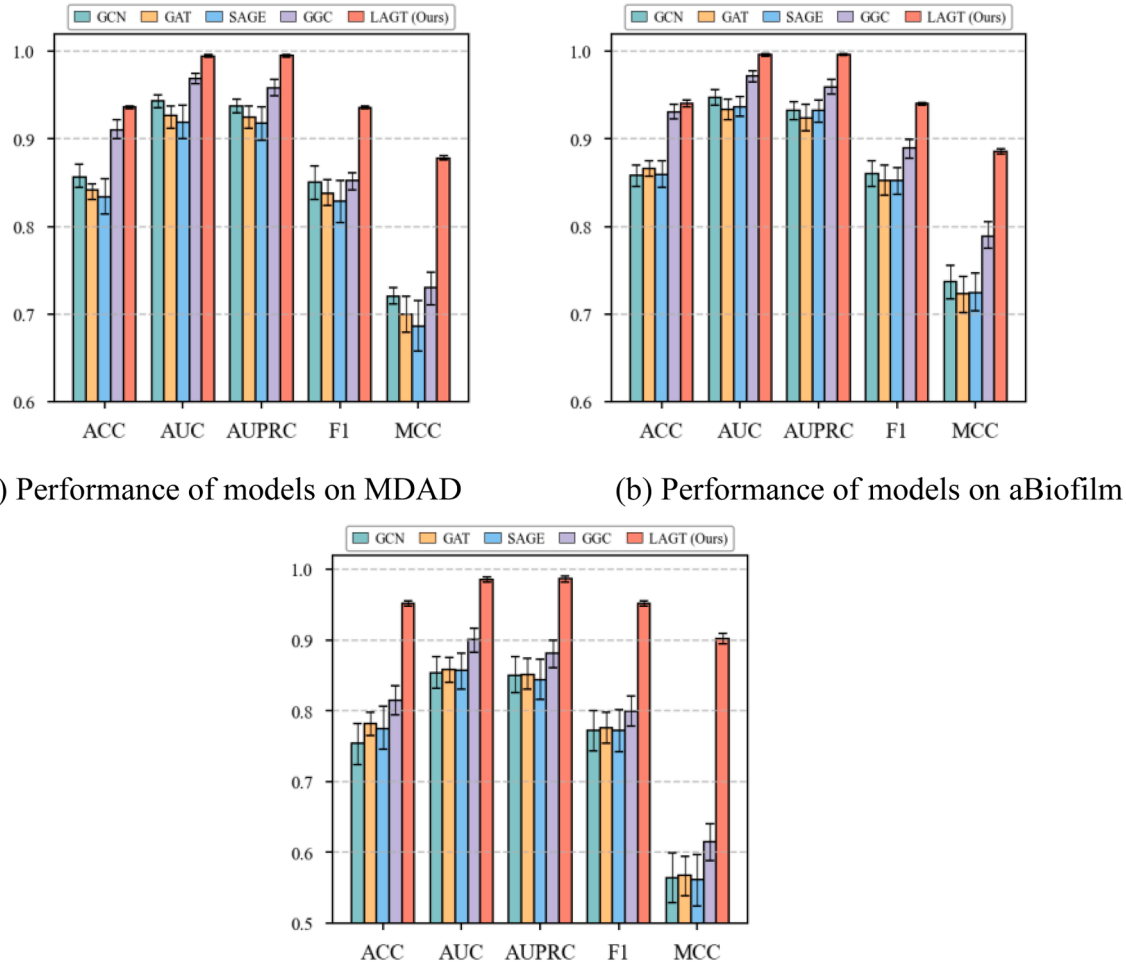
($p < 0.001$) and an AUPRC score of 6.2% ($p < 0.001$) on the MDAD dataset, dropped by an AUC score of 5.3% ($p < 0.001$) and an AUPRC score of 6.3% ($p < 0.001$) on the aBiofilm dataset, and dropped by an AUC score of 6.3% ($p < 0.001$) and an AUPRC score of 7.2% ($p < 0.001$) on the DrugVirus dataset. The outcomes underscore the necessity of integrating feature pathways, while highlighting the comparatively important contribution of deep structural representations captured by the LAGT module.

The specific contribution of the reciprocal distillation module (RDM) was isolated for drug-microbe association prediction by evaluating the HGCRL w/o RDM variant. By retaining the semantic and structural pathways but deactivating the alignment loss, the performance of HGCRL still showed a noticeable decrease, which exhibits an AUC decrease of 1.5% ($p < 0.01$) and AUPRC decrease of 1.1% ($p < 0.01$) on the MDAD dataset, an AUC decrease of 1.2% ($p < 0.01$) and AUPRC decrease of 1.1% ($p < 0.01$) on the aBiofilm dataset, and an AUC decrease of 2.0% ($p < 0.01$) and AUPRC decrease of 2.1% ($p < 0.01$) on the DrugVirus dataset. The results further confirm that the RDM is not merely a passive conduit for feature learning, but plays a crucial role in synergistically enhancing and integrating the representations from the two views. Moreover, the semantic and structural pathways can align effectively to capture a comprehensive and robust predictive signal across datasets.

The integration of LAGT, GCNConv, and RDM addresses key challenges in microbe-drug association prediction, such as sparse and heterogeneous graph structures, imbalanced modalities, and noisy semantic features in datasets. LAGT (structural pathway) combines positional information extraction, edge/global attention and layer-adaptive aggregation to capture multi-scale dependencies, enhancing representation distinguishability. GCNConv (semantic pathway) uses graph convolutional layers to propagate and aggregate node-specific semantic features, effectively capturing latent similarities while mitigating noise through neighborhood smoothing. RDM employs asymmetric distillation with EMA, stop-gradient, and regularization for stable cross-modal transfer.

2) Effect of the topological structure feature encoder

The section validates the impact of the topological structure feature encoder on HGCRL. Fig. 5 shows the performance variations of the HGCRL model with diverse modules on the MDAD, aBiofilm, and DrugVirus datasets. We systematically replaced the layer-adaptive graph transformer(LAGT) of the HGCRL model with four widely-used GNN modules, including graph convolutional networks (GCN) [43], graph attention networks (GAT) [39], gated graph convolution (GGC) [56], and GraphSAGE (SAGE) [40] while keeping all other model components unchanged. For ACC, AUC, AUPRC, F1, and MCC, the HGCRL model achieves the best remarkable results with the LAGT-based encoder on the MDAD, aBiofilm, and DrugVirus datasets. Consequently, for the drug-related microbe prediction task, the proposed LAGT module is an efficient topological structure feature encoder. The remarkable results underscore the effectiveness of LAGT for topological information, and also highlights that the dynamic, multi-scale feature learning capabilities of our model are highly effective for dealing with complex



(a) Performance of models on MDAD

(b) Performance of models on aBiofilm

(c) Performance of models on DrugVirus

Fig. 5. Performance of HGCRL with different encoders on the MDAD, aBiofilm, and DrugVirus datasets.

microbe-drug graph data.

3) Analysis of Key Components within the LAGT Encoder

To evaluate the superiority of the LAGT-based encoder, we performed a finer-grained ablation study to dissect and validate the contributions of its core internal components. Specifically, we created three variants of the full HGCRL model by removing the initial Positional Encoding module, removing the final Layer-adaptive aggregation mechanism, and replacing the scaled dot-product attention fusion (SDAF) module with a standard vanilla attention mechanism. The comprehensive performance metrics for these variants, alongside the complete LAGT configuration, are summarized in Table 2. Notably, removing the positional encoding module had the most pronounced impact. Its removal led to AUC decreases of 4.8% ($p < 0.001$, 95% CI: 0.9360-0.9564) and AUPRC decreases of 6.1% ($p < 0.001$, 95% CI: 0.9178-0.9487) on the MDAD dataset, 5.3% ($p < 0.001$, 95% CI: 0.9323-0.9550) and 6.6% ($p < 0.001$, 95% CI: 0.9043-0.9572) on the aBiofilm dataset, and 11.8% ($p < 0.001$, 95% CI: 0.9808-0.9845) and 14.1% ($p < 0.01$, 95% CI: 0.9791-0.9864) on the DrugVirus dataset. Replacing the SDAF with vanilla attention also caused noticeable drops, such as AUC decreases of 3.3% ($p < 0.001$, 95% CI: 0.9515-0.9714) and AUPRC ($p < 0.01$, 95% CI: 0.9593-0.9773) decreases of 2.6% on MDAD, up to 5.9% ($p < 0.01$, 95% CI: 0.9031-0.9726) and 4.2% ($p < 0.01$, 95% CI: 0.9298-0.9792) on aBiofilm, and minor 0.3% (p not significant, 95% CI: 0.9808-0.9845) and 0.4% (p not significant, 95% CI: 0.9791-0.9864) on DrugVirus. Similarly, removing the layer-adaptive aggregation yielded consistent degradations, with average AUC and AUPRC decreases of 2.0%–2.6% across datasets.

The observations highlight that the robust performance arises from the harmonious integration of architectural innovations. The positional encoding imparts essential structural priors to enhance node distinguishability, and the SDAF adeptly models intricate relational dependencies. Thus, the layer-adaptive aggregation optimally synthesizes multi-scale representations.

4) Effect of the meta-path-based semantic feature encoder

The section validates the impact of the meta-path-based semantic feature encoder on HGCRL on the MDAD, aBiofilm, and DrugVirus datasets. Table 3 shows the performance variations of the HGCRL model with diverse modules on the MDAD, aBiofilm, and DrugVirus datasets. The GCN module of HGCRL was replaced with GAT, GraphSAGE, GGC, and the LAGT proposed by this paper, respectively. As presented in Table 3, we observed that the performance remained remarkably stable, indicating that increasing the complexity of the semantic encoder does not translate to better prediction outcomes. This is likely because the semantic feature encoder operates on curated meta-path-induced subgraphs, which are inherently more structured and less complex than the global heterogeneous biological graph. In the prediction task, the feature extraction of a standard GCN is sufficiently powerful to capture the relevant relational patterns between microbes and drugs. Consequently, we used GCN to design the meta-path-based semantic feature encoder, adhering to the principle of model parsimony. This decision not only enhances computational efficiency, but also reinforces that the key innovation of the work lies in the reciprocal distillation between distinct structural and semantic embeddings, rather than the individual complexity of the encoders.

5) Analysis of the special loss function and class imbalance

The section discussed and validated the impact of the positive-to-negative sample ratios based on the performance of HGCRL with focal loss and standard binary cross-entropy loss (BCE) loss. All experiments were conducted across three positive-to-negative sample ratios (1:1, 1:5, 1:10). Based on the MDAD, aBiofilm, and DrugVirus datasets, Fig. 6 shows a performance comparison of HGCRL with the Focal and BCE loss under various ratios of positive and negative samples. From Fig. 6, when the ratio of positive-to-negative samples is 1:1, the HGCRL model performs excellently in AUC and AUPRC under balanced conditions, indicating the ability of HGCRL to effectively extract informative features between positive and negative samples. Through the analysis of the special loss function and class imbalance, selecting an appropriate positive-to-negative sample ratio can enhance the prediction capability and adaptability of HGCRL across various situations. When dealing with a scarcity of positive samples, the oversampling or other balancing methods may improve the drug-related microbe prediction performance, and further enhance the effectiveness and reliability in practical applications.

Besides, HGCRL with focal loss consistently outperformed HGCRL with BCE loss in AUPRC scores across all scenarios. On the MDAD dataset, HGCRL with focal loss achieved AUPRC improvements of 1.1% (1:1), 4.5% (1:5), and 8.0% (1:10) over HGCRL with BCE Loss. The aBiofilm dataset demonstrated corresponding gains of 1.0%, 4.2%, and 2.1%, respectively. The DrugVirus dataset exhibited the most substantial improvements: 6.2% at the 1:1 ratio, 22.7% at the 1:5 ratio, and 16.4% at the 1:10 ratio. Similar performance improvement of HGCRL is observed for AUC metrics on three datasets. These results demonstrate that the advantages of focal loss become more pronounced as class imbalance intensifies. The dynamic sample weighting mechanism effectively down-weights well-classified samples of drug-microbe associations, enabling the model to focus on challenging samples of drug-microbe associations. This successfully mitigates the adverse effects of skewed data distributions and enhances the model's capacity to identify authentic microbe-drug associations in imbalanced datasets.

4.4. Parameter analysis

This section assessed the robustness and adaptability of HGCRL, and

Table 2
Performance of HGCRL variants with key LAGT components.

Datasets	Variant	ACC	AUC	AUPRC	F1	MCC
MDAD						
w/o Layer-Adaptive		0.8816±0.0084	0.9764±0.0050	0.9781±0.0045	0.9131±0.0083	0.8417±0.0110
w/o Pos. Encoding		0.8704±0.1130	0.9462±0.0856	0.9332±0.1055	0.8104±0.1453	0.6620±0.1762
Vanilla Attention		0.8828±0.1085	0.9615±0.0847	0.9683±0.0805	0.8812±0.1089	0.7817±0.1436
Full LAGT		**0.9363±0.0026**	**0.9943±0.0010**	**0.9946±0.0009**	**0.9358±0.0024**	**0.8782±0.0043**
aBiofilm						
w/o Layer-Adaptive		0.9136±0.0981	0.9704±0.0720	0.9783±0.0551	0.9125±0.0989	0.8371±0.1307
w/o Pos. Encoding		0.8882±0.0803	0.9437±0.0904	0.9307±0.1380	0.8369±0.1223	0.7015±0.1559
Vanilla Attention		0.8625±0.2135	0.9378±0.1580	0.9545±0.1333	0.8616±0.2115	0.7366±0.3145
Full LAGT		**0.9402±0.0021**	**0.9963±0.0011**	**0.9965±0.0009**	**0.9399±0.0021**	**0.8857±0.0041**
DrugVirus						
w/o Layer-Adaptive		0.8897±0.0041	0.9593±0.0035	0.9623±0.0040	0.8864±0.0044	0.7863±0.0088
w/o Pos. Encoding		0.7752±0.0031	0.8689±0.0164	0.8468±0.0260	0.7553±0.0234	0.5567±0.0360
Vanilla Attention		0.9466±0.0054	0.9826±0.0013	0.9827±0.0026	0.9467±0.0055	0.8933±0.0109
Full LAGT		**0.9512±0.0050**	**0.9854±0.0026**	**0.9862±0.0033**	**0.9512±0.0050**	**0.9025±0.0100**

Table 3
Performance of different GNN encoders for meta-path-based semantic feature encoder.

Datasets	GNN	ACC	AUC	AUPRC	F1	MCC
MDAD	GCN	0.9363±0.0026	**0.9943±0.0010**	**0.9946±0.0009**	0.9358±0.0024	0.8782±0.0043
	GAT	0.9347±0.0039	0.9924±0.0021	0.9931±0.0016	0.9346±0.0024	0.8758±0.0028
	SAGE	0.9348±0.0044	0.9930±0.0006	0.9936±0.0003	0.9351±0.0044	0.8769±0.0078
	GGC	**0.9378±0.0032**	0.9929±0.0017	0.9936±0.0016	**0.9373±0.0018**	**0.8808±0.0010**
	LAGT	0.9319±0.0038	0.9931±0.0008	0.9936±0.0008	0.9312±0.0037	0.8697±0.0066
aBiofilm	GCN	0.9402±0.0021	**0.9963±0.0011**	**0.9965±0.0009**	0.9399±0.0021	0.8857±0.0041
	GAT	0.9370±0.0029	0.9954±0.0011	0.9957±0.0013	0.9364±0.0029	0.8794±0.0039
	SAGE	0.9379±0.0044	0.9952±0.0007	0.9955±0.0006	0.9374±0.0049	0.8811±0.0089
	GGC	0.9379±0.0047	0.9947±0.0020	0.9950±0.0016	0.9375±0.0024	0.8811±0.0022
	LAGT	**0.9412±0.0052**	0.9954±0.0009	0.9957±0.0008	**0.9410±0.0054**	**0.8876±0.0099**
DrugVirus	GCN	**0.9512±0.0050**	**0.9854±0.0026**	**0.9862±0.0033**	**0.9512±0.0050**	**0.9025±0.0100**
	GAT	0.9431±0.0063	0.9835±0.0027	0.9839±0.0038	0.9429±0.0061	0.8859±0.0120
	SAGE	0.9226±0.0146	0.9745±0.0036	0.9755±0.0051	0.9237±0.0139	0.8473±0.0185
	GGC	0.9325±0.0081	0.9783±0.0042	0.9787±0.0049	0.9330±0.0051	0.8671±0.0064
	LAGT	0.9208±0.0146	0.9744±0.0057	0.9754±0.0061	0.9194±0.0158	0.8397±0.0156

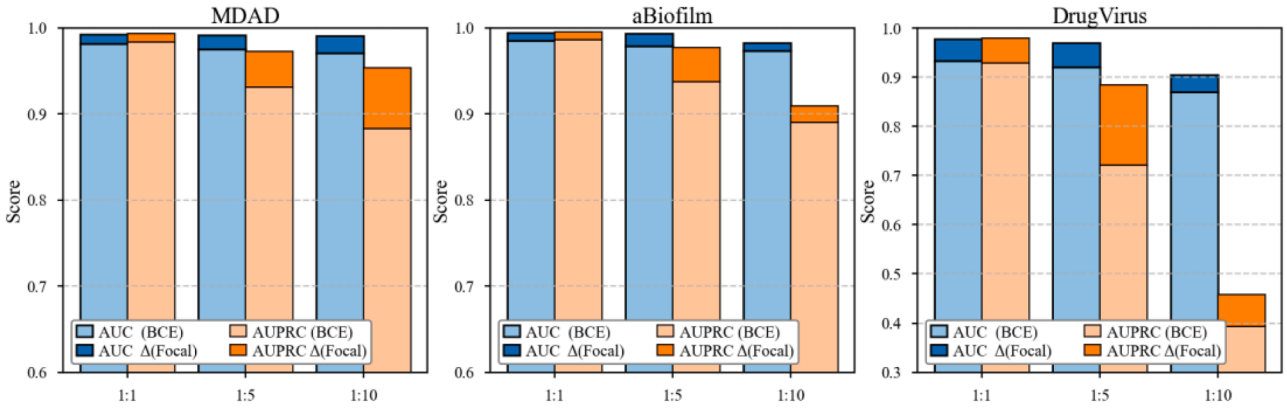


Fig. 6. Comparison of the focal and BCE loss under various imbalance ratios of positive and negative samples.

all experiments were performed under a 5-fold cross-validation. The parameters mainly include the number of attention heads, the weight of the reciprocal distillation module(λ_{ctr}), the focal loss parameters γ and α, and the key parameters within the RDM module(γ_1, γ_2 and μ_{var}, ν_{cov}).

1) Impact of attention heads

First, we analyzed the number of attention heads in the multi-head attention mechanism within the range [1,4,8,16]. Fig. 7 shows the results of HGCRL with the number of attention heads, and indicates that the model performance is optimal when attention heads are 4. The multi-head attention allows the model to capture diverse feature representations from different subspaces, which is beneficial for modeling complex microbe–drug associations. However, too few heads (e.g., 1) limit the representation capacity, while too many heads (e.g., 8 or 16) may introduce redundant information and noise, increase parameter complexity, and thereby degrade performance. Thus, an appropriate number of heads, such as 4, achieves a good balance between expressiveness and stability.

2) Impact of the overall distillation weight

In this section, we evaluated the performance of HGCRL with the weight λ_{ctr} of RDM. Specifically, we tested the performance of HGCRL with different values in the range [0.1, 0.2, 0.3, 0.4, 0.5]. Fig. 8 shows that the HGCRL model achieved the best performance at $\lambda_{ctr} = 0.3$. A lower weight (e.g., 0.1 or 0.2) is insufficient for the model to fully benefit from the self-supervised representations, making it harder to enrich feature embeddings. Conversely, a higher weight (e.g., 0.4 or 0.5) overemphasizes the auxiliary distillation objective, which interferes

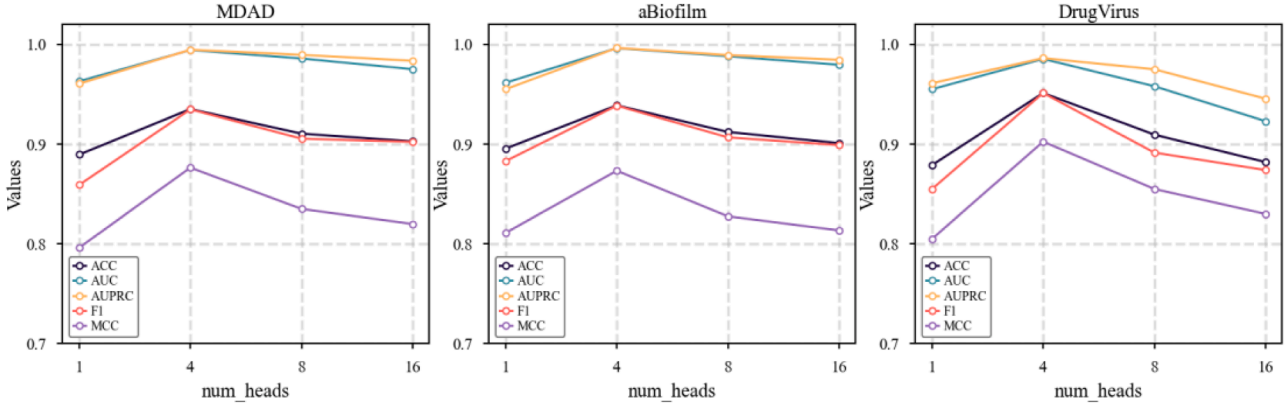


Fig. 7. Impact of the number of attention heads of HGCRL on microbe-drug association prediction.

with the primary prediction task and leads to performance degradation. Therefore, setting $\lambda_{ctr} = 0.3$ provides a proper balance between leveraging self-supervised signals and preserving task-specific optimization, resulting in more stable and accurate predictions.

3) Analysis of focal loss parameters

The experimental analysis was conducted based on parameters γ and α for the focal loss part, which regulates the weight between easy/hard and positive/negative samples, respectively. We conducted a grid search and constructed the parameter heatmap shown in Fig. 9 based on AUC score. The heatmap reveals that the model achieved optimal performance when $\gamma = 2.0$ and $\alpha = 1.0$. This setting allows the model to emphasize hard-to-classify samples without excessively down-weighting easier ones, while maintaining a balanced treatment of positive and negative samples. In particular, $\gamma = 2.0$ effectively reduces the dominance of easily classified pairs, forcing the model to focus on more informative and challenging cases, and $\alpha = 1.0$ prevents bias toward either class under imbalanced conditions. Therefore, this parameter combination mitigates class imbalance, enhances robustness, and ensures that the learned embeddings of microbe–drug pairs are both discriminative and interpretable.

4) Balancing alignment and regularization in RDM

For the reciprocal distillation module, we investigated the interplay between the weights γ_{align} and γ_{reg}. The parameters balance the cross-modal alignment loss $\mathscr{L}_{align}$ and the latent space regularization loss $\mathscr{L}_{reg}$, respectively. The heatmap analysis in Fig. 10 shows that performance is optimal for microbe-drug association prediction when γ_{align} is approximately 1.5 and γ_{reg} is 0.1, indicating the primary importance of the cross-modal alignment objective. This setting highlights that effective prediction relies more on aligning heterogeneous microbe and drug representations into a common latent space, while only a small amount of regularization is sufficient to maintain representation stability. If γ_{align} is too low, the embeddings of microbes and drugs cannot be well aligned, limiting predictive capacity. On the other hand, an overly large γ_{reg} may constrain the latent space excessively and suppress useful information. Therefore, the chosen balance ensures that embeddings remain discriminative, interpretable, and generalizable across different datasets.

5) Analysis of variance and covariance regularization weights

For regularization parameters (μ_{var} and ν_{cov}) in the internal RDM, a grid search was also performed on the MDAD, aBiofilm, and DrugVirus datasets. As illustrated in the heatmap in Fig. 11, the results show that the best performance is achieved with $\mu_{var} = 30.0$ and $\nu_{cov} = 1.0$. This setting strikes a proper balance in structuring the shared latent space. $\mu_{var} = 30.0$ ensures sufficient variance across embedding dimensions, which prevents feature collapse and improves the expressiveness of representations, while $\nu_{cov} = 1.0$ effectively decorrelates feature dimensions, reducing redundancy and overfitting. If the variance weight is too small, embeddings may become degenerate and lack diversity. If too large, they may introduce noise and instability. Similarly, excessive covariance penalization could overly constrain the latent space and harm predictive performance. Thus, the parameter combination provides the stability and discriminability, and the embeddings of microbe–drug pairs are interpretable and well-suited for association prediction.

4.5. Case studies

While the pathogenic mechanisms of microbes in response to drugs remain incompletely understood, existing studies indicate their partial involvement in disease treatment processes. Given the complexity and time-intensive nature of elucidating drug mechanisms, computational methods can accelerate the development of new therapeutic strategies. To assess HGCRL's performance in realistic scenarios, we conducted case studies on Curcumin and Epigallocatechin gallate (EGCG).

Curcumin, a natural polyphenol from turmeric rhizomes, is renowned for its antioxidant, anti-inflammatory, and antimicrobial properties [57]. It inhibits microbial biofilm formation and quorum sensing, as shown in studies on Pseudomonas aeruginosa [58] and Candida albicans, where it also enhances sensitivity to antifungal drugs [59]. Despite poor bioavailability, its broad-spectrum activity positions it as a promising candidate for combination therapies. Fig. 12(a) presents the UMAP visualization of drug-microbe associations centred on Curcumin. Curcumin clusters with predicted microbial targets: Serratia marcescens, Proteus mirabilis, and Actinomyces oris. This proximity in embedding space, driven by shared high-dimensional features learned by the model, suggests common molecular characteristics—such as membrane structures or metabolic pathways—making them susceptible to Curcumin. Confirmed interactions exist for Serratia marcescens and Proteus mirabilis (PMID: 31180643 and PMID: 25914690), while Actinomyces oris is a novel prediction with similar features.

EGCG, the primary catechin in green tea, exhibits potent antioxidant and antimicrobial effects [60]. It synergizes with β-lactam antibiotics against multidrug-resistant Staphylococcus aureus by inhibiting β-lactamase [61] and disrupts S. aureus growth, toxin production, and cell membranes via gene transcription changes [62]. Fig. 12(b) shows the UMAP visualization for EGCG, with tight clustering of Candida albicans, Vibrio vulnificus, Vibrio cholerae, and Serratia marcescens. This spatial arrangement indicates shared vulnerabilities, potentially in cell membrane integrity or oxidative stress responses, based on the model's learned embeddings. Literature supports interactions with Candida albicans and Vibrio cholerae (PMID: 26263544 and PMID: 2381042), while Vibrio vulnificus is a novel target with comparable characteristics.

In case studies, the model was first trained on all known and previously unseen associations between microbes and drugs, and then, the candidate microbes related to drugs were ranked in descending order based on the predicted scores. During the training, drug-microbe

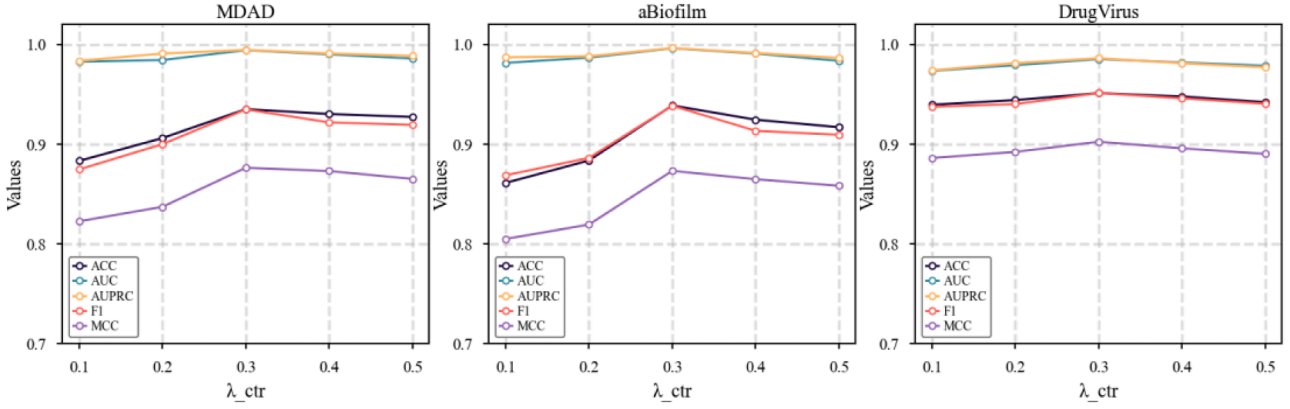


Fig. 8. Impact of the weight λ_{ctr} in RDM on the performance of HGCRL.

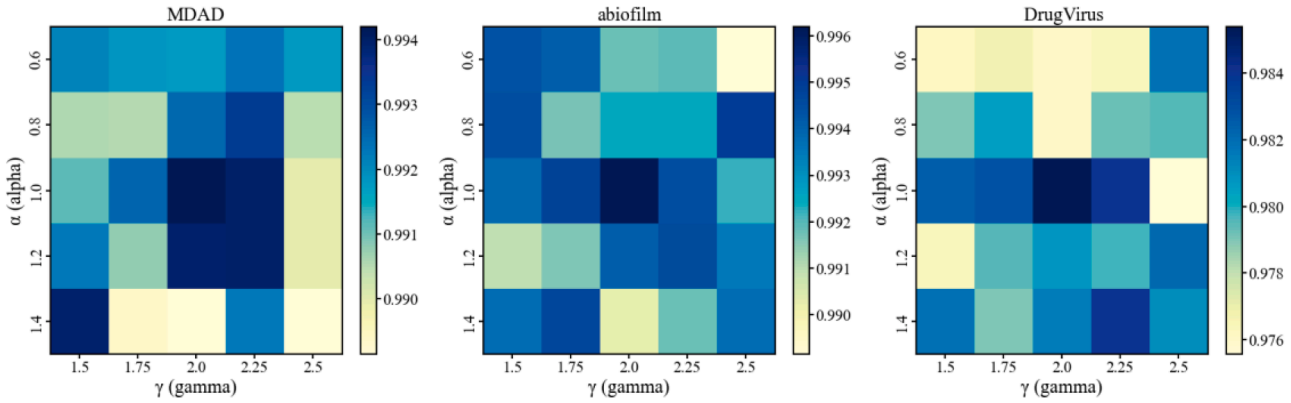


Fig. 9. Parameter analysis of γ and α in the focal loss of HGCRL on the MDAD, aBiofilm, and DrugVirus datasets.

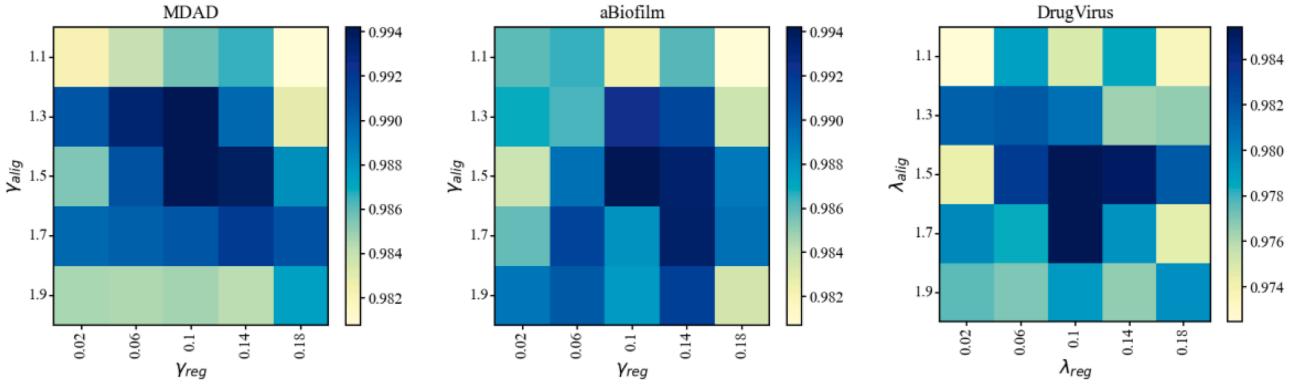


Fig. 10. Parameter analysis of RDM on the MDAD, aBiofilm, and DrugVirus datasets.

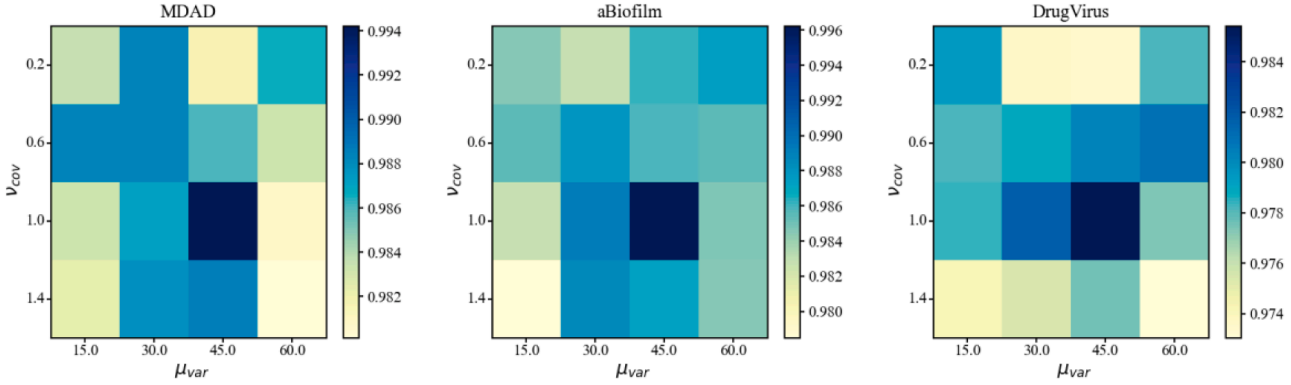
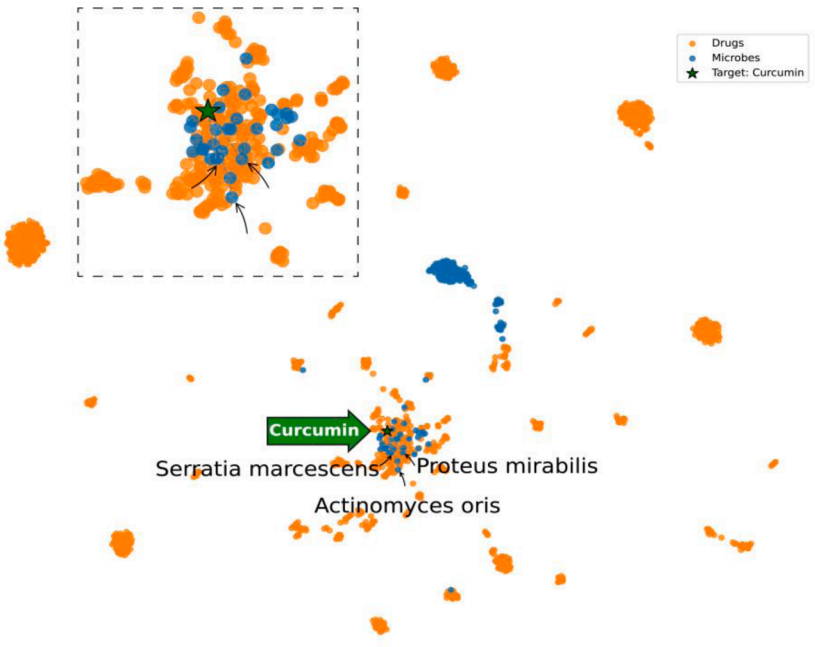


Fig. 11. Parameter analysis of RDM on the MDAD, aBiofilm, and DrugVirus datasets.

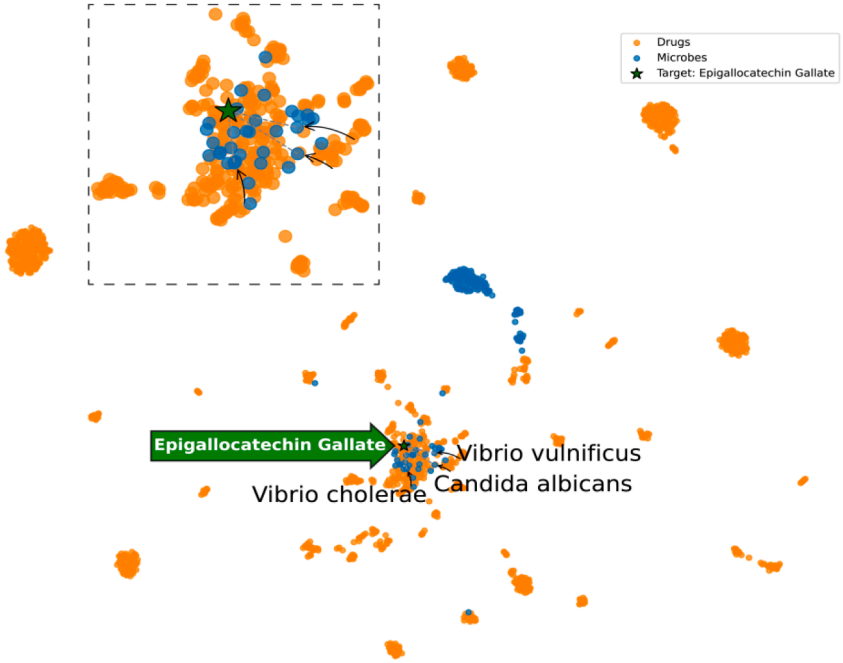
associations with Curcumin and Epigallocatechin gallate were intentionally omitted from the training dataset. In the inference, the model is used to predict and rank microbes and microbes based on the potential interactions with Curcumin, Epigallocatechin gallate, respectively, using prediction scores. The findings were then validated, as shown in Tables 4 and 5. The validation confirmed that 18 potential microbes potentially interacting with Curcumin were identified through the biomedical literature from PubMed(https://pubmed.ncbi.nlm.nih.gov/). For Epigallocatechin Gallate, the literature from PubMed confirmed 17 potential microbes with potential interactions. Although the microbes have not been confirmed to interact with Curcumin and Epigallocatechin Gallate in the literature from PubMed, the experiments indicate the following results: 1) Obesumbacterium proteus and Actinomyces oris changes may affect the Therapeutic effects of Curcumin; 2) the Therapeutic effects of Epigallocatechin Gallate may be affected by Vibrio vulnificus, Streptococcus epidermidis, and Staphylococcus cohnii. The above results demonstrate that the HGCRL model could effectively predict potential drug-related microbes, highlighting the practical value.

5. Conclusion

Accurate microbe-drug association prediction is essential for drug development and precision medicine, yet existing methods struggle with complex biological graph topologies and data imbalance. In this work, we developed an efficient computational model combining layer-adaptive graph transformer with reciprocal distillation to extract global and local features, while avoiding explicit negative sampling for



(a) Case study for Curcumin.



(b) Case study for Epigallocatechin Gallate (EGCG).

Fig. 12. Case study validation of predicted drug-microbe associations using UMAP.

robust learning from sparse data. Extensive experiments demonstrated superior performance compared to state-of-the-art methods, and our model accurately predicts potential drug-related microbes and could contribute to early disease detection and precision medicine applications.

Future work will focus on two key directions. First, incorporating physicochemical properties (molecular descriptors, drug solubility, and microbe metabolic profiles) will capture drug-microbe interaction mechanisms at the molecular level, improving predictions for novel drugs with limited experimental data. Second, applying the model to large-scale biological networks will enable discovery of rare microbe-drug associations across diverse microbial ecosystems. However, scaling to other biomedical tasks presents challenges including computational scalability for multi-million node networks, domain-specific feature engineering requirements, and maintaining model interpretability across heterogeneous biological data types. These developments will advance microbiome-informed precision medicine.

CRediT authorship contribution statement

Yanbu Guo: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization.

Table 4
The top 20 Microorganisms responsive to Curcumin.

Microbe name	Evidence	Microbe name	Evidence
Serratia marcescens	PMID: 31180643	Escherichia coli	PMID: 35326785
Pseudomonas aeruginosa	PMID: 32421995	Staphylococcus aureus	PMID: 27956904
Bacillus cereus	PMID: 31180643	Klebsiella pneumoniae	PMID: 35322691
Burkholderia pseudomallei	PMID: 32421995	Actinomyces oris	Unconfirmed
Proteus mirabilis	PMID: 25914690	Streptococcus mutans	PMID: 23778072
Vibrio cholerae	PMID: 21808656	Candida spp.	PMID: 30108820
Candida albicans	PMID: 33592455	Aeromonas hydrophila	PMID: 32947883
Vibrio vulnificus	PMID: 22092562	Obesumbacterium proteus	Unconfirmed
Staphylococcus epidermidis	PMID: 24773066	Vibrio harveyi	PMID: 23354447
Salmonella enterica	PMID: 20634977	Bacillus subtilis	PMID: 35456852

Table 5
The top 20 microorganisms responsive to Epigallocatechin Gallate.

Microbe name	Evidence	Microbe name	Evidence
Candida albicans	PMID: 26263544	Vibrio harveyi	PMID: 26849416
Burkholderia cenocepacia	PMID: 30737168	Enterococcus faecalis	PMID: 25526623
Staphylococcus aureus	PMID: 32679222	Burkholderia pseudomallei	PMID: 33346350
Streptococcus pneumoniae	PMID: 28402019	Pseudomonas aeruginosa	PMID: 34066609
Vibrio cholerae	PMID: 2381042	Stenotrophomonas maltophilia	PMID: 15980368
Staphylococcus epidermidis	PMID: 28661772	Dengue Virus	PMID: 28744282
Klebsiella pneumoniae	PMID: 21596543	Streptococcus epidermidis	Unconfirmed
Streptococcus mutans	PMID: 34525984	Staphylococcus cohnii	Unconfirmed
Vibrio vulnificus	Unconfirmed	Candida parapsilosis	PMID: 16831406
Candida tropicalis	PMID: 24366745	Serratia marcescens	PMID: 12841932

Quanming Guo: Writing – review & editing, Visualization, Software, Investigation, Data curation. **Shengli Song:** Validation, Supervision, Resources, Methodology, Formal analysis, Conceptualization. **Yihan Wang:** Visualization, Validation, Software, Investigation. **Jinde Cao:** Resources, Methodology, Investigation.

Declaration of competing interest

We declare that we have no conflict of interest.

Acknowledgements

The authors thank the editors and anonymous reviewers for constructive comments. This work was supported by the National Natural Science Foundation of China (62403437, 62573122 , 62576098 and 62573176), the Project of Science and Technology in Henan Province (252102211070, 252102211066 and 252102211069), the Key Program of Higher Education of Henan Province (25A120003) and the Key Research and Development Program of Shaanxi (2024GX-YBXM545).

Data availability

Data will be made available on request.

References

[1] R. Adhikari, S. Das, S. Das, Neural networks-based framework for recognizing streaming patterns in magnetized Maxwell–Oldroyd-B blood blended with tetra-hybrid nanoparticles and microbes over stenosis in an elastic artery, Eng. Appl. Artif. Intell. 136 (2024) 109048.
[2] X. Du, J. Li, B. Wang, J. Zhang, T. Wang, J. Wang, NRGCNMDA: microbe-drug association prediction based on residual graph convolutional networks and conditional random fields, Interdiscip. Sci. Comput. Life Sci. 17 (2025) 344–358.
[3] H. Yu, H. Xu, X. Yang, Z. Zhang, J. Hu, J. Lu, et al., Gut microbiota-based pharmacokinetic-pharmacodynamic study and molecular mechanism of specnuezhenide in the treatment of colorectal cancer targeting carboxylesterase, J. Pharm. Anal. 13 (2023) 1024–1040.
[4] Z. Huang, Z. Xiao, C. Ao, L. Guan, L. Yu, Computational approaches for predicting drug-disease associations: a comprehensive review, Front. Comput. Sci. 19 (2025) 195909.
[5] Y. Long, J. Luo, Association mining to identify microbe drug interactions based on heterogeneous network embedding representation, IEEE J. Biomed. Health Inform. 25 (2021) 266–275.
[6] L. Deng, Y. Huang, X. Liu, H. Liu, Graph2MDA: a multi-modal variational graph embedding model for predicting microbe–drug associations, Bioinformatics 38 (2022) 1118–1125.
[7] H. Huang, Y. Sun, M. Lan, H. Zhang, G. Xie, GNAEMDA: microbe-drug associations prediction on graph normalized convolutional network, IEEE J. Biomed. Health Inform. 27 (2023) 1635–1643.
[8] Z. Tian, Y. Yu, H. Fang, W. Xie, M. Guo, Predicting microbe–drug associations with structure-enhanced contrastive learning and self-paced negative sampling strategy, Brief Bioinform. 24 (2023) bbac634.
[9] Y. Long, M. Wu, Y. Liu, C.K. Kwoh, J. Luo, X. Li, Ensembling graph attention networks for human microbe–drug association prediction, Bioinformatics 36 (2020) i779–ii86.
[10] Y. Tan, J. Zou, L. Kuang, X. Wang, B. Zeng, Z. Zhang, et al., GSAMDA: a computational model for predicting potential microbe–drug associations based on graph attention network and sparse autoencoder, BMC Bioinform. 23 (2022) 492.
[11] Y. Ma, Q. Liu, Generalized matrix factorization based on weighted hypergraph learning for microbe-drug association prediction, Comput. Biol. Med. 145 (2022) 105503.
[12] Q. Ma, Y. Tan, L. Wang, GACNNMDA: a computational model for predicting potential human microbe-drug associations based on graph attention network and CNN-based classifier, BMC Bioinform. 24 (2023) 35.
[13] Z. Li, X. Li, X. Tang, Y. Wang, AMCF-RDP: a self-attention-based multi-source and cascade framework for the identification of drug–protein relationships, Mol. Divers. (2025) 1–19.
[14] D. Chen, T. Zhang, H. Cui, J. Gu, P. Xuan, KNDM: a knowledge graph transformer and node category sensitive contrastive learning model for drug and microbe association prediction, J. Chem. Inf. Model. 65 (2025) 4714–4728.
[15] J. Lai, Z. Zhang, B. Zeng, L. Wang, GTFKAN: a novel microbe-drug association prediction model based on Graph Transformer and Fourier Kolmogorov-Arnold Networks, J. Mol. Biol. (2025) 169201.
[16] L. Barbadilla-Martínez, N. Klaassen, B. van Steensel, J. de Ridder, Predicting gene expression from DNA sequence using deep learning models, Nat. Rev. Genet. 26 (2025) 666–680.
[17] M.E. Consens, C. Dufault, M. Wainberg, D. Forster, M. Karimzadeh, H. Goodarzi, et al., Transformers and genome language models, Nat. Mach. Intell. 7 (2025) 346–362.
[18] Y. Guo, H. Shen, W. Li, C. Li, C. Jin, Deep effective k-mer representation learning for polyadenylation signal prediction via co-occurrence embedding, Knowl. Based Syst. 256 (2022) 109887.
[19] Y. Guo, D. Pang, M. Huang, D. Zhou, J. Cao, Deep dual dynamic gated convolutional model with interpretable pooling for genomic signal prediction, Data Intell. (2025) 1–21, https://doi.org/10.3724/2096-7004.di.2025.0051.
[20] W. Li, Y. Guo, B. Wang, B. Yang, Learning spatiotemporal embedding with gated convolutional recurrent networks for translation initiation site prediction, Pattern Recognit. 136 (2023) 109234.
[21] Y. Guo, Y. Xin, J. Cao, Y. Xu, D. Zhou, Deep gate information bottleneck-based prediction model for complex disease-related micro-ribonucleic acids via heterogeneous biological networks, Eng. Appl. Artif. Intell. 156 (2025) 111222.
[22] J. Xia, L. Wu, G. Wang, J. Chen, S.Z. Li, ProGCL: rethinking hard negative mining in graph contrastive learning, editors, in: C. Kamalika, J. Stefanie, S. Le, S. Csaba, N. Gang, S. Sivan (Eds.), Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research: PMLR, 2022, pp. 24332–24346.
[23] R. Zhang, Z. Wang, X. Wang, Z. Meng, W. Cui, Mhtan-dti: metapath-based hierarchical transformer and attention network for drug–target interaction prediction, Brief Bioinform. 24 (2023) bbad079.
[24] X. Qiu, H. Wang, X. Tan, Z. Fang, GK BertDTA: a graph representation learning and semantic embedding-based framework for drug-target affinity prediction, Comput. Biol. Med. 173 (2024) 108376.

[25] Y. Xia, X. Pan, H.B. Shen, Heterogeneous sampled subgraph neural networks with knowledge distillation to enhance double-blind compound-protein interaction prediction, Structure 32 (2024) 611–620, e4.
[26] X. Wang, Q. Zhao, J. Wang, FedKD-CPI: combining the federated knowledge distillation technique to accomplish synergistic compound-protein interaction prediction, Methods 234 (2025) 275–283.
[27] B. Wang, T. Wang, X. Du, J. Li, J. Wang, P. Wu, Microbe-drug association prediction model based on graph convolution and attention networks, Sci. Rep. 14 (2024) 22327.
[28] L. Fan, L. Wang, X. Zhu, A novel microbe-drug association prediction model based on stacked autoencoder with multi-head attention mechanism, Sci. Rep. 13 (2023) 7396.
[29] B. Wang, W. Zhao, X. Du, J. Zhang, C. Zhang, L. Wang, et al., Predicting potential microbe-disease associations based on heterogeneous graph attention network and deep sparse autoencoder, Eng. Appl. Artif. Intell. 147 (2025) 110301.
[30] H. Li, H. Zhen-Jie, Z. Wen-Guang, J. Qu, Y. Hai-Bin, Y. Chen, Prediction of potential drug-microbe associations based on matrix factorization and a three-layer heterogeneous network, Comput. Biol. Chem. 104 (2023) 107857.
[31] A. Ullah, Y. Meng, J.F.F. Mendes, OVED-rank: a ranking scheme to evaluate complex network spreaders' influence through the concept of effective distance and orbital velocity, Inf. Process. Manag. 62 (2025) 104201.
[32] A. Ullah, S.U. Din, N. Khan, C.B. Mawuli, J. Shao, Towards investigating influencers in complex social networks using electric potential concept from a centrality perspective, Inform. Fusion 109 (2024) 102439.
[33] B. Li, H. Wang, X. Tan, Q. Li, J. Chen, X. Qiu, Adaptive heterogeneous graph reasoning for relational understanding in interconnected systems, J. Supercomput. 81 (2025) 112.
[34] Y.Z. Sun, D.H. Zhang, S.B. Cai, Z. Ming, J.Q. Li, X. Chen, MDAD: a special resource for microbe-drug associations, Front. Cell. Infect. Microbiol. 8 (2018) 424.
[35] A. Rajput, A. Thakur, S. Sharma, M. Kumar, aBiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance, Nucleic Acids Res. 46 (2018) D894–D900.
[36] P.I. Andersen, A. Ianevski, H. Lysvand, A. Vitkauskiene, V. Oksenych, M. Bjørås, et al., Discovery and development of safe-in-man broad-spectrum antiviral agents, Int. J. Infect. Dis. 93 (2020) 268–276.
[37] X. Fu, J. Zhang, Z. Meng, K.I. Magnn, Metapath aggregated graph neural network for heterogeneous graph embedding, in: Proceedings of the Web Conference 2020, 2020, pp. 2331–2341.
[38] X. Wu, A. Ajorlou, Z. Wu, A. Jadbabaie, Demystifying oversmoothing in attention-based graph neural networks, Adv. Neural Inf. Process. Syst. 36 (2023) 35084–35106.
[39] Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph attention networks. arXiv preprint arXiv:171010903. 2017.
[40] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Process. Syst. 30 (2017).
[41] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, X. Sun, Measuring and relieving the over-smoothing problem for graph neural networks from the topological view, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 3438–3445.
[42] Oono K., Suzuki T. Graph neural networks exponentially lose expressive power for node classification. arXiv preprint arXiv:190510947. 2019.
[43] Kipf T.N., Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:160902907. 2016.
[44] X. Chen, K. He, Exploring simple siamese representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021, 2021, pp. 15750–15758.
[45] Zhang C., Zhang K., Yoo C.D., Kweon I.S. How does simsiam avoid collapse without negative samples? a unified understanding with self-supervised contrastive learning. arXiv preprint arXiv:2203.16262.
[46] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017).
[47] C. Wu, J. Pfrommer, M. Zhou, J. Beyerer, Self-supervised generative-contrastive learning of multi-modal euclidean input for 3d shape latent representations: a dynamic switching approach, IEEE Trans. Multimed. 26 (2023) 8432–8441.
[48] Weis M.A., Hansel L., Lüddecke T., Ecker A.S. Self-supervised graph representation learning for neuronal morphologies. arXiv preprint arXiv:211212482. 2021.
[49] Thakoor S., Tallec C., Azar M.G., Azabou M., Dyer E.L., Munos R., et al. Large-scale representation learning on graphs via bootstrapping. arXiv preprint arXiv: 210206514. 2021.
[50] Bardes A., Ponce J., LeCun Y. VICReg: variance-invariance-covariance regularization for self-supervised learning. 2022.
[51] T.Y. Ross, G. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2980–2988.
[52] J. Gu, T. Zhang, Y. Gao, S. Chen, Y. Zhang, H. Cui, et al., Neighborhood topology-aware knowledge graph learning and microbial preference inferring for drug-microbe association prediction, J. Chem. Inf. Model. 65 (2025) 435–445.
[53] Z. Tian, Y. Yu, F. Ni, Q. Zou, Drug-target interaction prediction with collaborative contrastive learning and adaptive self-paced sampling strategy, BMC Biol. 22 (2024) 216.
[54] H. Kuang, Z. Zhang, B. Zeng, X. Liu, H. Zuo, X. Xu, et al., A novel microbe-drug association prediction model based on graph attention networks and bilayer random forest, BMC Bioinform. 25 (2024) 78.
[55] J. Zhao, L. Kuang, A. Hu, Q. Zhang, D. Yang, C. Wang, OGNNMDA: a computational model for microbe-drug association prediction based on ordered message-passing graph neural networks, Front. Genet. 15 (2024) 1370013.
[56] Li Y., Tarlow D., Brockschmidt M., Zemel R. Gated graph sequence neural networks. arXiv preprint arXiv:151105493. 2015.
[57] R.R. Kotha, D.L. Luthria, Curcumin: biological, pharmaceutical, nutraceutical, and analytical aspects, Molecules (2019) 24.
[58] A. Shariati, M. Noei, M. Askarinia, A. Khoshbayan, A. Farahani, Z. Chegini, Inhibitory effect of natural compounds on quorum sensing system in Pseudomonas aeruginosa: a helpful promise for managing biofilm community, Front. Pharmacol. 15 (2024) 1350391.
[59] S. Hajifathali, S. Lesan, E. Lotfali, E. Salimi-Sabour, M. Khatibi, Investigation of the antifungal effects of curcumin against nystatin-resistant Candida albicans, Dent. Res. J. 20 (2023) 50.
[60] S. Legeay, M. Rodier, L. Fillon, S. Faure, N. Clere, Epigallocatechin Gallate: a review of its beneficial properties to prevent metabolic syndrome, Nutrients 7 (2015) 5443–5468.
[61] W.H. Zhao, N. Asano, Z.Q. Hu, T. Shimamura, Restoration of antibacterial activity of beta-lactams by epigallocatechin gallate against beta-lactamase-producing species depending on location of beta-lactamase, J. Pharm. Pharmacol. 55 (2003) 735–740.
[62] A. Kitichalermkiat, M. Katsuki, J. Sato, T. Sonoda, Y. Masuda, K.I. Honjoh, et al., Effect of epigallocatechin gallate on gene expression of Staphylococcus aureus, J. Glob. Antimicrob. Resist. 22 (2020) 854–859.