

# Drug Repositioning via Multi-View Representation Learning With Heterogeneous Graph Neural Network

Li Peng , Cheng Yang , Jiahuai Yang, Yuan Tu, Qingchun Yu, Zejun Li, Min Chen , and Wei Liang 

**Abstract**—Exploring simple and efficient computational methods for drug repositioning has emerged as a popular and compelling topic in the realm of comprehensive drug development. The crux of this technology lies in identifying potential drug-disease associations, which can effectively mitigate the burdens caused by the exorbitant costs and lengthy periods of conventional drugs development. However, existing computational drug repositioning methods continue to encounter challenges in accurately predicting associations between drugs and diseases. In this paper, we propose a Multi-view Representation Learning method (MRLHGNN) with Heterogeneous Graph Neural Network for drug repositioning. This method is based on a collection of data from multiple biological entities associated with drugs or diseases. It consists of a view-specific feature aggregation module with meta-paths and auto multi-view fusion encoder. To better utilize local structural and semantic information from specific views in heterogeneous graph, MRLHGNN employs a feature aggregation model with variable-length meta-paths to expand the local receptive field. Additionally, it utilizes a transformer based semantic aggregation module to aggregate semantic features across different view-specific graphs. Finally, potential drug-disease associations are obtained through a multi-view fusion decoder with an attention mechanism. Cross-validation experiments demonstrate the effectiveness and interpretability of the MRLHGNN in comparison to nine state-of-the-art approaches. Case studies further reveal that MRLHGNN can serve as a powerful tool for drug repositioning.

Received 11 July 2023; revised 7 October 2023, 28 November 2023, 7 February 2024, and 22 July 2024; accepted 23 July 2024. Date of publication 29 July 2024; date of current version 7 March 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62172158, in part by the National Natural Science Foundation of Hunan Province under Grant 2023JJ30264 and Grant 2024JJ7115, and in part by the Scientific Research Project of Hunan Education Department under Grant 22A0350. (*Corresponding authors:* Li Peng; Wei Liang.)

Li Peng and Wei Liang are with the School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China, and also with the Hunan Key Laboratory for Service computing and Novel Software Technology Xiangtan 411201, China (e-mail: plpeng@hnu.edu.cn; wliang@hnust.edu.cn).

Cheng Yang, Jiahuai Yang, Yuan Tu, and Qingchun Yu are with the School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China (e-mail: yangchengyjs@163.com; jason1325070309@163.com; ty09060418@163.com; yqc@hnust.edu.cn).

Zezun Li and Min Chen are with the School of Computer Science and Engineering, Hunan Institute of Technology, Hengyang 421002, China (e-mail: lzjfox@hnit.edu.cn; chenmin@hnit.edu.cn).

Digital Object Identifier 10.1109/JBHI.2024.3434439

**Index Terms**—Drug, disease, drug repositioning, multi-view representation learning, heterogeneous graph neural network, meta-path.

## I. INTRODUCTION

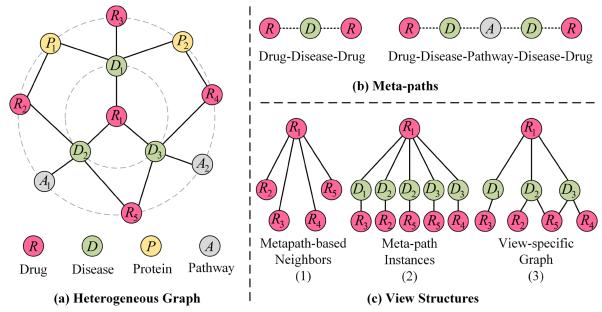
THE development of symptomatic drugs for an emerging disease is a high-risk, extremely expensive and time-consuming process with a low success rate [1]. Drug repositioning can apply existing drugs or compounds to new therapeutic targets, and also provide an efficient screening method for drug combination strategies to treat diseases [2], [3]. This approach is particularly effective in treating complex diseases because it increases the likelihood of identifying effective therapeutic drug combinations [4], [5]. In addition, the method provides a reliable candidate for the design of targeted drugs, thereby shortening the drug development cycle. Revealing potential drug-disease associations is a crucial step in the process of drug repositioning, which aims to identify new indications for existing drugs. It reduces the cost of unnecessary biological experiments and enhances the success rate of drug development. With the explosive growth of medical data and the advancement of deep learning techniques, various data-driven computational models have been proposed specifically for predicting drug small molecule-related markers [6], [7], [8], [9], disease-related markers [10], [11], [12]. These methodologies have significant implications for drug repositioning, and some can even be directly applied. The primary objective of drug repositioning approaches is to generate predictions for unknown drug-disease pairs or reconstruct drug-disease association matrix based on representations of drug and disease feature. Existing predictive approaches for drug repositioning can be broadly classified into two categories: classical machine learning methods and deep learning techniques.

Traditional approaches primarily focus on matrix completion and network propagation. Zhang et al. [13] used known associations to construct similarity measure graphs for diseases and drugs, and then derived potential drug-disease association scores using label propagation. Yang et al. [14] constructed an heterogeneous association matrix based on known association and similarities, and subsequently utilized the bounded kernel norm approach (BNNR) to derive association scores of potential drug-disease associations. Afterwards, yang et

al. [15] proposed HGIMC, a heterogeneous graph inference method based on matrix complementation (HGIMC), to predict new drug indications using heterogeneous association information. Ji et al. [16] developed DTINet, a method for predicting new drug-target interactions through label propagation and induction matrix completion based on network similarity information.

Currently, several new deep learning-based heterogeneous graph neural network approaches are being proposed to model drug-disease associations, offering high flexibility and scalability. Their performance results are significantly more competitive than traditional approaches. For instance, LAGCN [17] is Layer Attention Graph Convolutional Network that utilizes layer attention mechanisms to learn representations of drug or disease nodes in a heterogeneous drug-disease network. DRHGCN [18] is an information fusion-based graph convolutional network designed with inter- and intra-domain embeddings for learning representations of drug or disease nodes. DRWBNCF [19] is a neural collaborative filtering approach for drug repositioning based on neighbourhood interactions. The three aforementioned methods focus solely on predicting missing values in the drug-disease bipartite association network, overlooking related biological entities and their rich semantic context. REDDA [20] merges three attention mechanisms to learn drug/disease representations from the node embedding block, topological subnet embedding block, graph attention block and layer attention block of a heterogeneous graph convolutional network in a sequential process. However, REDDA integrates enhanced biological node data but limits its graph convolution to first-order neighbors, neglecting multi-order semantic insights and local structure of heterogeneous graph. Based on heterogeneous graph networks, Gu et al. [21] proposed a multi-instance learning method (MilGNet) for drug repositioning. MilGNet learns node feature representations at the meta-path level using a pseudo-meta-path instance generator and bidirectional translational embedding projector. MilGNet primarily focuses on meta-path level feature aggregation, neglecting the view level. And, its meta-path instance enumeration significantly increases computational memory usage as the graph's entity count grows. Additionally, multi-view representation learning [22], [23], [24], which aggregates node features through meta-paths of drugs' view-specific graphs and diseases' view-specific graphs, and has not yet been examined in the field of drug repositioning.

To address these issues, we proposed a Multi-view Representation Learning with Heterogeneous Graph Neural Network, called MRLHGNN, for drug repositioning. MRLHGNN consists of two main modules that are innovative in the drug repositioning: a module for multi-view representation learning, which explores the intrinsic heterogeneity between nodes, and an auto multi-view fusion decoder for predicting drug-disease associations. After conducting a series of comprehensive experiments, it has been conclusively proven that MRLHGNN outperforms 9 state-of-the-art methods on the benchmark dataset and has potential for discovering new drug-disease associations. In summary, the aims and contributions of this work can be summarized as follows:



**Fig. 1.** An illustrative example of a heterogeneous graph and some key concepts. (a) The local structure of node  $R_1$  in the heterogeneous graph. (b) Two metapaths in heterogeneous graph. (c) Three types of metapath-based local structures of drug node  $R_1$ .

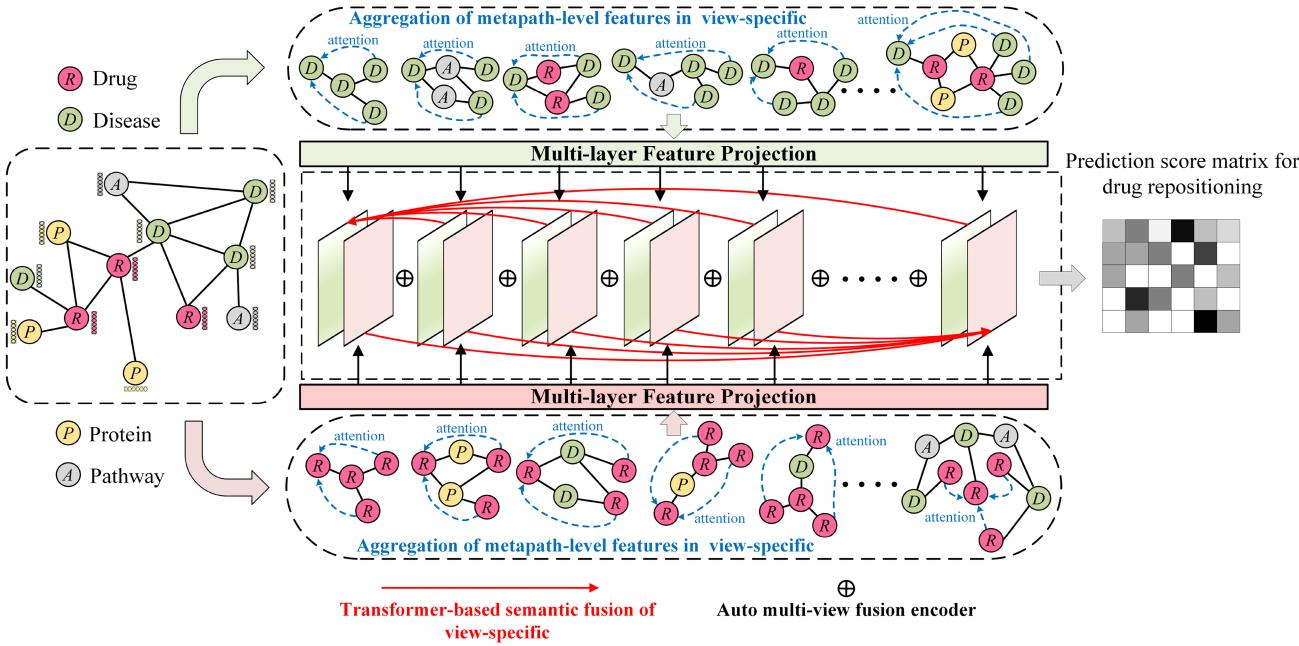
- To enhance the characterization of drug and disease nodes in the heterogeneous graph, we incorporated other biological entities (such as proteins and pathways) that are associated with a particular drug or disease, resulting in a more locally informative heterogeneous graph.
- This work proposed a view-specific feature aggregation module, including metapath-level node feature aggregation with attention-based and view-level node feature fusion with transformer-based, to leverage the inherent heterogeneity information of local structures in the specific view.
- We employed an automatic multi-view fusion encoder to generate the final drug-disease association matrix, facilitating more accurate drug repositioning. Moreover, case studies have affirmed the efficacy of the MRLHGNN approach.

## II. PROBLEM DEFINITION

**Definition 1 (Heterogeneous Graph):** A heterogeneous graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T}_v, \mathcal{T}_e)$  is made up of a vertex set  $v_i \in \mathcal{V}$  and an edge set  $e_{i,j} \in \mathcal{E}$ , along with node type mapping function  $\mathcal{M}_v(v_i) : \mathcal{V} \rightarrow \mathcal{T}_{v_i}$  and edge type mapping function  $\mathcal{M}_e(e_{i,j}) : \mathcal{E} \rightarrow \mathcal{T}_{e_{i,j}}$ .  $\mathcal{O}$  and  $\mathcal{Q}$  denote the predefined sets of object types and edge types, respectively, where  $|\mathcal{O}| + |\mathcal{Q}| > 2$ . An example of this is given in Fig. 1(a).

**Definition 2 (Metapath):** Consider  $O_i \in \mathcal{O}$  and  $Q_i \in \mathcal{Q}$  denote a node type and an edge type, respectively. A metapath starting at node type  $O_1$  and ending at node type  $O_l$  can be expressed as  $\mathcal{P} \triangleq O_1 \xrightarrow{Q_1} O_2 \xrightarrow{Q_2} \dots \xrightarrow{Q_{l-2}} O_{l-1} \xrightarrow{Q_{l-1}} O_l$ , which describes a composite relation  $\mathcal{Q} = Q_1 \circ Q_2 \circ \dots \circ Q_{l-1}$  between node types  $Q_1$  and  $Q_{l-1}$ , where  $\circ$  denotes the composition operator over relations. Examples are given in Fig. 1(b).

**Definition 3 (Metapath-based Graph):** Given a metapath  $\mathcal{P}_v : O_1 \rightarrow O_2 \rightarrow \dots \rightarrow O_N$  and a target node  $v$  with edge type  $Q_{l-1}$ , the view-specific graph  $\mathcal{GP}_v$  is a directed graph evoked from metapath-based neighbourhoods and intermediate nodes following the metapath along with  $v$  itself. An illustrative example is shown in Fig. 1(c).



**Fig. 2.** The architecture of MRLHGNN.

### III. MATERIALS AND METHODS

In this section, the experimental benchmark dataset, the construction of heterogeneous graph, and the theory of the MRL-HGNN model are introduced. The architecture of MRLHGNN is shown in Fig. 2.

#### A. Data Information

We integrated the Fdataset [25], Cdataset [26] and additional data downloaded from KEGG [27] and CTD [28] as our drug-disease association data, which contained 894 drugs (r), 454 diseases (d) and 2704 identified associations (d-r, r-d). The matrix  $Y(r_i, d_j) \in \mathbb{R}^{N_r \times N_d}$  represents drug-disease associations matrix, with  $N_r$  being 894,  $N_d$  being 454. We collected 454 disease- or 894 drug-related biologic relationship data from DrugBank [29], CTD [28], KEGG [27], STRING [30] and UniProt [31], obtaining 18878 proteins, 314 pathways, 1048575 protein-protein associations (p-p) [30], 1669 pathway-pathway associations (a-a) [27], 4397 drug-protein associations (r-p, p-r) [29], and 19530 disease-pathway associations (d-a, a-d) [27].

For drug-drug associations (r-r), we calculated binary molecular fingerprints [32] of simplified molecular input line-entry system (SMILES) that represent the structure of each drug. Then, we used the Tanimoto method to calculate the drug-drug pair similarity matrix  $S_r$ . To simplify the construction of heterogeneous graphs, we converted the drug-drug similarity to a binarized value. Specifically, in drug similarity matrix  $S_r$ , we filtered the  $topk$  similarity values for each drug to convert to 1 and the others to convert to 0. After the conversion, the  $S'_r$  is procured as the ultimate drug-drug associations.

To calculate the disease-disease associations (d-d), we used the Medical Subject Headings (MeSH) identifiers to obtain the disease semantic similarity matrix  $S_d$  through directed acyclic

**TABLE I**  
THE DETAILS OF BENCHMARK DATASET

Data type	Number	Resource
Drug	894	DrugBank
Disease	454	KEGG, CTD
Protein	18 878	KEGG, UniProt
Pathway	314	KEGG, CTD
Drug-Disease	2704	KEGG
Drug-Protein	4397	DrugBank
Disease-Pathway	19 530	CTD
Protein-Protein	1 048 575	STRING
Pathway-Pathway	1669	KEGG
Drug-Drug	-	Sim_Comput
Disease-Disease	-	Sim_Comput

<sup>1</sup> Note: Sim\_Comput means that the associations are yielded using the threshold  $topk$  of top similarity values.

graph (DAG) [33]. Like the drug similarity conversion process, we also used the  $topk$  process to convert the similarities  $S_d$  to disease-disease associations  $S'_d$ . The baseline dataset used in this work is shown in Table I. Furthermore, we expanded the relevant data from the publicly available dataset used in the studies by Yu et al. [17] and Guo et al. [34]. This creates new Dataset-B and Dataset-C, which are used to test the sensitivity of model performance to the associated data. The details of Dataset-B and Dataset-C are shown in Table II.

#### B. Construction of Drug-Disease Heterogeneous Graph

The drug repositioning, utilizing graph neural networks can be defined as a link prediction task. It aims to predict the probability of whether a potential association exists between given drug ( $r$ ) and a disease ( $d$ ) node. Hence, assume a heterogeneous graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  where  $\mathcal{V} = (\mathcal{V}_r, \mathcal{V}_d, \mathcal{V}_p, \mathcal{V}_a)$  is the node set

**TABLE II**  
THE DETAILS OF DATASET-B AND DATASET-C

Data type	Dataset-B		Dataset-C	
	Number	Resource	Number	Resource
Drug	269	Yu et al. [17]	1025	Guo et al. [34]
Disease	598	Yu et al. [17]	2062	Guo et al. [34]
Protein	6040	Uniprot	1649	STRING
Pathway	295	KEGG, CTD	491	KEGG, CTD
Drug-Disease	18416	Yu et al. [17]	18416	Guo et al. [34]
Drug-Protein	2107	DrugBank	11107	DrugBank
Disease-Pathway	9656	CTD	165230	CTD
Protein-Protein	592926	STRING	19237	STRING
Pathway-Pathway	1504	KEGG	1606	KEGG
Drug-Drug	-	Sim_Comput	-	Sim_Comput
Disease-Disease	-	Sim_Comput	-	Sim_Comput

<sup>1</sup> Note: Sim\_Comput means that the associations are yielded using the threshold *topk* of top similarity values.

and  $\mathcal{E} = (\mathcal{E}_{(r-r)}, \mathcal{E}_{(d-d)}, \mathcal{E}_{(p-p)}, \mathcal{E}_{(a-a)}, \mathcal{E}_{(r-d, d-r)}, \mathcal{E}_{(r-p, p-r)}, \mathcal{E}_{(d-a, a-d)})$  is the edge set. the associations matrix  $Y$  of heterogeneous graph  $\mathcal{G}$  can be defined as:

$$Y = \begin{bmatrix} S'_r & Y_{(r,p)} & 0 & Y_{(r,d)} \\ Y_{(r,p)}^T & Y_{(p,p)} & 0 & 0 \\ 0 & 0 & Y_{(a,a)} & Y_{(a,d)} \\ Y_{(r,d)}^T & 0 & Y_{(a,d)}^T & S'_d \end{bmatrix}. \quad (1)$$

Then, the heterogeneous graph  $\mathcal{G}$  is constructed using Deep Graph Library [35]. Additionally, we utilized the calculated similarities  $S_r$  and  $S_d$  to represent the chemical structural features of the drug and the semantic features of the disease. the node feature matrix  $X \in \mathbb{R}^{(N_r+N_d) \times (N_r+N_d)}$  of drug and disease in graph  $\mathcal{G}$  can be initialized and represented as:

$$X = \begin{bmatrix} S_r & 0 \\ 0 & S_d \end{bmatrix}. \quad (2)$$

The details of the constitutive heterogeneous graph in MRL-HGNN are described in Supplementary Material.

### C. Aggregation of Metapath-Level Features in View-Specific Graph

To capture the heterogeneity information between nodes in the local structure, we treat each semantic (represented by the metapath)  $\mathcal{P}_v$  as a view and perform view-specific metapath-level message aggregation to obtain the node feature embedding for each view. Specifically, according to the semantic information of heterogeneity between nodes, the original heterogeneous graph is decomposed into multiple view-specific graph [23]. This operation has the advantage of decoupling multiple semantics. Moreover, compared to neighborhoods and meta-path instances based on meta-paths, the view-specific graph formed by drug or disease nodes through intermediary neighboring nodes retains a more comprehensive view-centric local structure under each meta-path in the heterogeneous semantics. As a consequence, under semantic  $\mathcal{P}_v$ , the view-specific graph  $\mathcal{GP}_v$  preserves the original and complete multi-hop structure in the heterogeneous graph. An example is shown in Fig. 1(c).

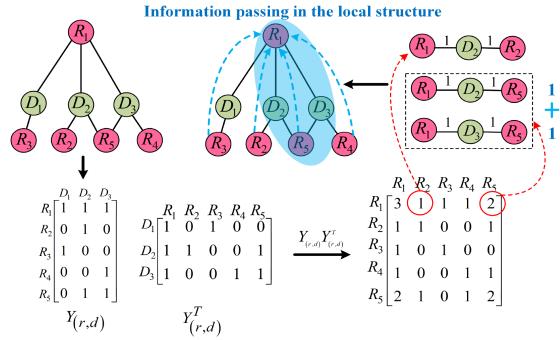


Fig. 3. Information aggregation in the local structure.

To effectively process the feature information of drug nodes, we develop a meta-path based feature-weighted aggregator, similar to disease node. This approach offers the advantage of aggregating node features guided by view-specific graph. Furthermore, incorporating variable-length meta-path instances into a view-specific graph can enhance the receptive field of local node feature aggregation. A detailed example is shown in Fig. 3. And, drug- and disease-specific view graph of the MRLHGNN model is depicted in Table III. The aggregation process can be expressed as:

$$u_i = \left\{ z_i^{\mathcal{GP}_r} = \left( \frac{1}{|S^{\mathcal{GP}_r}|} \sum_{p_r(i,j) \in S^{\mathcal{GP}_r}} w_j x_j \right) : \mathcal{GP}_r \in \Phi_r \right\}, \quad (3)$$

where  $u_i \in U_r$  and  $U_r = \{X_{\mathcal{GP}_r} : \mathcal{GP}_r \in \Phi_r\}$  is a list of different semantic feature matrices generated for the set  $\Phi_r$  of all given view-specific  $\mathcal{GP}_r$  of drug.  $S^{\mathcal{GP}_r}$  is the set of all metapath instances which correspond to view-specific graph  $\mathcal{GP}_r$ ,  $|S^{\mathcal{GP}_r}|$  denotes the number of meta-path instances in  $\mathcal{GP}_r$ , and  $p_r(i,j)$  is one metapath instance containing target node  $i$  and source node  $j$ .  $w_j$  is the weight of attention given to feature  $x_j \in X[0 : N_r] \in \mathbb{R}^{N_r \times (N_r+N_d)}$  of neighbouring nodes on the meta-path. In executing the same feature aggregation process, the weighted feature aggregator also generated a list  $U_d = \{X_{\mathcal{GP}_d} : \mathcal{GP}_d \in \Phi_d\}$  which contains different semantic feature matrices that correspond to different view-specific graph  $\mathcal{GP}_d$  of disease, respectively. The process of node feature aggregation at the meta-path level in view-specific graph can be clearly observed in Fig. 2. The step of aggregating node features on metapaths enumerates all metapath-based neighbors for each metapath in view-specific graph, which imposes a high computational expense on the model when the number of metapath instances grows exponentially with the length of the metapath. Therefore, inspired by Sun et al. [36], we proposed a new approach using adjacency matrix multiplication to further simplify (3). The simplification proceeds as follows:

$$X_{\mathcal{GP}_r} = \hat{Y}_{(r,e_1)} \hat{Y}_{(e_1,e_2)} \cdots \hat{Y}_{(e_{l-1},r)} (w \odot X_r), \quad (4)$$

where  $\mathcal{P}_r = e_1 e_2 \cdots e_{l-1} r$  is a  $l$ -hop metapath, and  $\hat{Y}_{(r,e_1)}$  is row-normalized form of adjacency matrix  $Y_{(r,e_1)}$  between node type  $r$  and  $e_1$  in a view-specific  $\mathcal{GP}_r$ .  $w \in \mathbb{R}^{N_r \times 1}$  denotes the feature aggregation weight vector of the drug neighbour nodes

**TABLE III**  
DESCRIPTION OF VIEW-SPECIFIC GRAPH IN MRLHGNN

Node type	View-specific graph set $\Phi_r$ of drug	Metapath for constructing view-specific graph of drug	Node type	View-specific graph set $\Phi_d$ of disease	Metapath for constructing view-specific graph of disease
drug	$\mathcal{GP}_r^1$	$R \rightarrow R$	disease	$\mathcal{GP}_d^1$	$D \rightarrow D$
	$\mathcal{GP}_r^2$	$R \rightarrow P \rightarrow R$		$\mathcal{GP}_d^2$	$D \rightarrow A \rightarrow D$
	$\mathcal{GP}_r^3$	$R \rightarrow D \rightarrow R$		$\mathcal{GP}_d^3$	$D \rightarrow R \rightarrow D$
	$\mathcal{GP}_r^4$	$R \rightarrow R \rightarrow P \rightarrow R$		$\mathcal{GP}_d^4$	$D \rightarrow D \rightarrow A \rightarrow D$
	$\mathcal{GP}_r^5$	$R \rightarrow R \rightarrow D \rightarrow R$		$\mathcal{GP}_d^5$	$D \rightarrow D \rightarrow R \rightarrow D$
	$\mathcal{GP}_r^6$	$R \rightarrow P \rightarrow R \rightarrow R$		$\mathcal{GP}_d^6$	$D \rightarrow A \rightarrow D \rightarrow D$
	$\mathcal{GP}_r^7$	$R \rightarrow D \rightarrow R \rightarrow R$		$\mathcal{GP}_d^7$	$D \rightarrow R \rightarrow D \rightarrow D$
	$\mathcal{GP}_r^8$	$R \rightarrow D \rightarrow A \rightarrow D \rightarrow R$		$\mathcal{GP}_d^8$	$D \rightarrow R \rightarrow P \rightarrow R \rightarrow D$

on the metapath  $\mathcal{P}_r$ .  $\odot$  denotes the element-wise product of vectors and  $X_r = X[0 : N_r] \in \mathbb{R}^{N_r \times (N_r + N_d)}$  is the raw feature matrix of all drug nodes. With the same simplified feature aggregation method, we can obtain disease node feature embeddings matrix  $X_{\mathcal{GP}_d}$  at the metapath level within the disease view-specific graph.

Feature embedding vectors with varying dimensions can pose challenges when aggregating view-level information in subsequent modules. To solve this problem, we employ view-specific feature transformations to convert heterogeneous feature embedding vectors into features with consistent dimensionality. Taking the example of the conversion of features for a specific-view of drug, the specific process of conversion is as follows:

$$H'_{\mathcal{GP}_r} = MLP_{\mathcal{GP}_r}(X_{\mathcal{GP}_r}), \quad (5)$$

where  $MLP_{\mathcal{GP}_r}()$  is a multi-layer perception block for specific-view  $\mathcal{GP}_r$  of drug, including a normalization layer, a nonlinear layer, and a dropout layer between two consecutive linear layers. Similar to the feature conversion process of drug view-specific, we also obtained feature transformation matrix  $H'_{\mathcal{GP}_d}$  of disease for each view-specific  $\mathcal{GP}_d$ .

#### D. Transformer-Based Semantic Feature Fusion of View-Specific

In this section, our aim is to fuse different representations from multiple view-specific graph. To produce embedding vectors with efficient representation learning capacity, we proposed a transformer-based [37] semantic fusion module for further exploring the cross-relationships between feature vectors after meta-path aggregation in different view-specific. Specifically, with the different view list  $\Phi_r = \{\mathcal{GP}_r^1, \mathcal{GP}_r^2, \dots, \mathcal{GP}_r^M\}$  of drug and converted semantic vectors  $\{h'_{\mathcal{GP}_r^1}, h'_{\mathcal{GP}_r^2}, \dots, h'_{\mathcal{GP}_r^M}\}$  for each drug node,  $M$  represents the number of view-specific graph for both drug and disease nodes. The transformer-based semantic feature fusion module for view-specific of drug learns the mutual attention for each pair of semantic vectors. For each semantic vector  $h'_{\mathcal{GP}_r^i}$ , it maps the vector into a query vector  $q_{\mathcal{GP}_r^i}$ , a key vector  $k_{\mathcal{GP}_r^i}$ , and a value vector  $v_{\mathcal{GP}_r^i}$ . The mutual attention weight  $\alpha_{(\mathcal{GP}_r^i, \mathcal{GP}_r^j)}$  is the dot product result of the query vector  $q_{\mathcal{GP}_r^i}$  and the key vector  $k_{\mathcal{GP}_r^j}$  after a softmax normalization. The output vector  $h_{\mathcal{GP}_r^i}$  of current semantic  $\mathcal{GP}_r^i$  is the weighted sum of all value vectors  $v_{\mathcal{GP}_r^j}$  plus a residual connection. The process of semantic fusion can be formulated as:

$$q_{\mathcal{GP}_r^i} = W_Q h'_{\mathcal{GP}_r^i}, \quad (6)$$

$$k_{\mathcal{GP}_r^i} = W_K h'_{\mathcal{GP}_r^i}, \quad (7)$$

$$v_{\mathcal{GP}_r^i} = W_V h'_{\mathcal{GP}_r^i}, \quad (8)$$

$$\alpha_{(\mathcal{GP}_r^i, \mathcal{GP}_r^j)} = \frac{\exp(q_{\mathcal{GP}_r^i} * k_{\mathcal{GP}_r^j}^T)}{\sum_{\mathcal{GP}_r^t \in \Phi_r} \exp(q_{\mathcal{GP}_r^i} * k_{\mathcal{GP}_r^t}^T)}, \quad (9)$$

$$h_{\mathcal{GP}_r^i} = \varphi \sum_{\mathcal{GP}_r^j \in \Phi_r} \alpha_{(\mathcal{GP}_r^i, \mathcal{GP}_r^j)} v_{\mathcal{GP}_r^j} + h'_{\mathcal{GP}_r^i}, \quad (10)$$

where  $W_Q$ ,  $W_K$ ,  $W_V$  and  $\varphi$  are trainable parameters shared by all view-specific of drug. Then, we obtained a list  $H_r = \{H_{\mathcal{GP}_r^1}, H_{\mathcal{GP}_r^2}, \dots, H_{\mathcal{GP}_r^M}\}$  of feature fusion matrices for each view-specific graph of drug. For the different view list  $\Phi_d = \{\mathcal{GP}_d^1, \mathcal{GP}_d^2, \dots, \mathcal{GP}_d^M\}$ , with the same transformer-based semantic feature fusion operation, we also obtained a list  $H_d = \{H_{\mathcal{GP}_d^1}, H_{\mathcal{GP}_d^2}, \dots, H_{\mathcal{GP}_d^M}\}$  of feature matrices, which contain matrices corresponding to different view-specific graph  $\mathcal{GP}_d$  in the set  $\Phi_d$ , respectively.

#### E. Auto Multi-View Fusion Decoder for Predicting Drug-Disease Associations

Inspired by yang et al. [38], to decode the semantic embedding vectors of drug and disease fused on different views to achieve the prediction mission, this section used an automatic multi-view fusion decoder to predict drug-disease associations. According to Table V, we performed one-to-one semantic decoding view of the semantic matrices of view-specific in the set  $H_r$  and  $H_d$ , respectively, and then integrated the decoded preliminary prediction score matrices using the attention mechanism to obtain the final drug-disease matrix. Furthermore, we also perform similar decoding of the initial drug node features  $H_0^r = X[0 : N_r]$  and the initial disease node features  $H_0^d = X[N_r : (N_r + N_d)]$  in the heterogeneous graph to obtain the initial view of the automatic decoder. The above process can be formulated as:

$$\begin{aligned} Y'_{(r,d)} &= \eta \left( H_0^r W_0 (H_0^d)^T \right) + \beta_{(\mathcal{GP}_r^1, \mathcal{GP}_d^1)} Y'_{(\mathcal{GP}_r^1, \mathcal{GP}_d^1)} + \dots \\ &\quad + \beta_{(\mathcal{GP}_r^M, \mathcal{GP}_d^M)} Y'_{(\mathcal{GP}_r^M, \mathcal{GP}_d^M)} \\ &= \eta \left( H_0^r W_0 (H_0^d)^T \right) + \sum_{n=1}^M \beta_{(\mathcal{GP}_r^n, \mathcal{GP}_d^n)} \\ &\quad \times \left( H_{\mathcal{GP}_r^n} W_n (H_{\mathcal{GP}_d^n})^T \right), \end{aligned} \quad (11)$$

where,  $\beta_{(\mathcal{GP}_r^n, \mathcal{GP}_d^n)}$  is the attention coefficient for each preliminary score matrix  $Y'_{(\mathcal{GP}_r^n, \mathcal{GP}_d^n)}$  and  $W_n$  is the parameter matrix

within each decoder.  $Y'_{(r,d)}$  is the final drug-disease prediction matrix.

### F. Optimization

We use a weighted cross-entropy loss function to balance the effect of an unbalanced dataset in order to optimize the parameters of the prediction model and ensure that MRLHGNN focus on confirmed drug-disease associations. For  $N_r$  drugs and  $N_d$  diseases in the heterogeneous network  $\mathcal{G}$ , with confirmed/unconfirmed drug-disease associations labeled  $S^+$  and  $S^-$ , respectively, the loss function of the MRLHGNN model can be expressed as:

$$\text{Loss} =$$

$$-\frac{1}{N_r + N_d} \left( \gamma \sum_{(r,d) \in S^+} \log Y'_{(r,d)} + \sum_{(r,d) \in S^-} (1 - \log Y'_{(r,d)}) \right), \quad (12)$$

where,  $\gamma = \frac{|S^-|}{|S^+|}$  is the balance weight,  $|S^+|$  and  $|S^-|$  are the number of confirmed/unconfirmed drug-disease associations in the training set.

## IV. EXPERIMENTS AND RESULTS

### A. Evaluation Metrics and Parameters Setting

To evaluate the overall performance of MRLHGNN, we use evaluation metrics such as Area Under the Receiver Operating Characteristic Curve (*AUC*), the Area Under the Precision-Recall curve (*AUPR*), *F1\_score*, *Accuracy*, *Recall*, *specificity*, and *Precision*. Among these metrics, *AUC*, *AUPR* and *F1\_score* are selected as core metrics to further describe the performance benefits of MRLHGNN. Our proposed MRLHGNN model uses the Adam optimizer to optimize the neural network parameters.

In the MRLHGNN model, the values of all hyperparameters are referred to the practice of previous researchers and finally determined by grid search, where the learning rate  $r$  is 0.005, the node feature dimension size  $k$  in the network layer is 128, the dropout rate  $dr$  is 0.4 and the setting of similarity threshold  $topk$  affecting the number of associations of disease-disease and drug-drug is set to 15. The details of the hyperparameters in MRLHGNN are described in Supplementary Material. In addition, all methods are compared under the same evaluation settings, which include the datasets we used and the similarity calculations for drug and disease. For the baseline model with publicly available code, we refer to the best hyper-parameters reported in its original paper to run the code.

All experiments are conducted on a Windows Pro PC with a GeForce RTX 3090 GPU, 32 GB of RAM (Random Access Memory) and Intel(R)Core(TM)i7-13700K CPU @ 5.40 GHz. All algorithms are implemented in PyTorch and compiled using Python 3.8.1. We have released the code on <https://github.com/biohnuster/MRLHGNN>.

### B. Baseline Methods

To evaluate the model performances of our proposed MRLHGNN, we compared it to nine baseline methods: (i) traditional machine learning methods including: NTSIM [13], BNNR [14], and HGIMC [15]; (ii) deep learning methods including: NIM-CGCN [39], LAGCN [17], DRHGCN [18], DRWBNCF [19], REDDA [20] and MilGNet [21].

### C. Comparison With Other Methods

Similar to [40] and [41], we employed five-fold cross-validation (5-CV) to evaluate the predictive performance of the methods. Specifically, we divided all positive samples (validated) and all negative samples (unvalidated) into five equal parts, respectively. Four equal parts of positive and negative samples are treated as the training set, and one equal part of positive and negative samples are considered as the test set. In particular, the final predictive evaluation results of the model are obtained by averaging the performance metrics value from 10 times 5-CV. The performance comparison are reported in Table IV, where MRLHGNN demonstrates a competitive advantage, apart from the *Specificity* and *Precision* metrics. The results in Table IV indicated that our proposed MRLHGNN outperforms the other nine baseline methods in terms of metrics including *AUC*, *AUPR*, *F1\_score*, and *Recall*. It shows relative improvements of 4.51%, 20.87%, 22.9%, and 32.5% respectively compared to suboptimal methods. For *Accuracy*, *Specificity* and *Precision* metrics, our method's performance remains comparable. Among the compared deep learning methods, there is a 13.47% improvement in *Precision* metric compared to the suboptimal method (DRHGCN). Benchmarking comparison results show that adding information on proteins, side effects into known drug-disease associations data while using a multi-view based mechanism can significantly improve the comprehensive prediction performance.

We also tested our model on two another datasets (Dataset-B and Dataset-C) to demonstrate the reliability of MRLHGNN on public datasets. The performance results of MRLHGNN and nine baseline methods are presented in Tables VI and VII, respectively. In Dataset-B, the MRLHGNN achieved the best performance among the compared baseline methods for all evaluation metrics except the *AUPR* (second-best performance advantage). Specifically, compared to the suboptimal methods, the MRLHGNN demonstrated improvements of 2.04% in *AUC*, 0.13% in *F1\_score*, 0.21% in *Accuracy*, 4.42% in *Recall*, 6.05% in *Specificity*, and 1.18% in *Precision*. In Dataset-C, the MRLHGNN achieved comparatively favorable performance. Apart from obtaining the third-best performance advantage in terms of *AUPR* and *F1\_score*, it outperformed the baseline methods in the remaining evaluation metrics, achieving either the best or second-best performance advantage. In these two additional datasets, statistical comparisons of evaluation metrics between the methods further highlight the relatively reliable predictive performance of the MRLHGNN.

To evaluate MRLHGNN's sensitivity to known drug-disease associations, we randomly removed a portion of

**TABLE IV**  
PERFORMANCE OF TEN METHODS IN TERMS OF *AUC*, *AUPR*, *F1\_score*, *Accuracy*, *Recall*, *Specificity*, AND *Precision* UNDER 5-CV ON BENCHMARK DATASET

Methods	<i>AUC</i>	<i>AUPR</i>	<i>F1_score</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Specificity</i>	<i>Precision</i>
NTSIM	$0.6700 \pm 0.0000$	$0.2550 \pm 0.0000$	$0.3960 \pm 0.0000$	$0.9950 \pm 0.0000$	$0.2480 \pm 0.0000$	$\underline{1.0000 \pm 0.0000}$	$0.9820 \pm 0.0000$
BNNR	$0.7180 \pm 0.0000$	$0.3110 \pm 0.0000$	$0.3970 \pm 0.0000$	$0.9950 \pm 0.0000$	$0.2500 \pm 0.0000$	$\underline{1.0000 \pm 0.0000}$	$0.9740 \pm 0.0000$
NIMCGCN	$0.7410 \pm 0.0020$	$0.1290 \pm 0.0060$	$0.2080 \pm 0.0008$	$0.9894 \pm 0.0008$	$0.2128 \pm 0.1234$	$0.9944 \pm 0.0008$	$0.1910 \pm 0.0486$
HGIMC	$0.7742 \pm 0.0040$	$0.2864 \pm 0.0030$	$0.3716 \pm 0.0081$	$0.9948 \pm 0.0004$	$0.2468 \pm 0.0068$	$\underline{0.9998 \pm 0.0004}$	$0.8110 \pm 0.0127$
LAGCN	$0.7876 \pm 0.0104$	$0.1720 \pm 0.0058$	$0.3734 \pm 0.0013$	$0.9512 \pm 0.0029$	$0.2950 \pm 0.0076$	$0.9548 \pm 0.0008$	$0.0418 \pm 0.0008$
MilGNet	$0.8100 \pm 0.0012$	$0.2167 \pm 0.0080$	$0.2943 \pm 0.0012$	$0.9909 \pm 0.0018$	$0.2851 \pm 0.0061$	$0.9956 \pm 0.0009$	$0.3041 \pm 0.0010$
DRWBNCF	$0.8408 \pm 0.0087$	$0.2134 \pm 0.0340$	$0.2754 \pm 0.0343$	$0.9906 \pm 0.0008$	$0.2740 \pm 0.0414$	$0.9952 \pm 0.0011$	$0.2804 \pm 0.0377$
REDDA	$0.9134 \pm 0.1100$	$0.2208 \pm 0.0495$	$0.2688 \pm 0.0393$	$0.9932 \pm 0.0044$	$0.2998 \pm 0.0402$	$0.9938 \pm 0.0016$	$0.2466 \pm 0.0474$
DRHGNC	$0.9118 \pm 0.0009$	$0.4558 \pm 0.0017$	$0.4910 \pm 0.0023$	$0.9940 \pm 0.0000$	$0.4568 \pm 0.0151$	$0.9972 \pm 0.0004$	$0.5326 \pm 0.0260$
MRLHGNN	<b><math>0.9585 \pm 0.0025</math></b>	<b><math>0.6645 \pm 0.0118</math></b>	<b><math>0.7200 \pm 0.0004</math></b>	<b><math>0.9959 \pm 0.0030</math></b>	<b><math>0.7818 \pm 0.0008</math></b>	$0.9974 \pm 0.0005$	$0.6673 \pm 0.0010$

<sup>1</sup> Note: The best results are marked in bold and the second-best results are marked as underlined.  
(mean  $\pm$  std).

**TABLE V**  
THE DETAILS OF AUTO MULTI-VIEW FUSION DECODER

Views in auto fusion decoder	Description	Attention coefficient for each views
View0	$H_0^T W_0 (H_0^d)^T$	$\eta$
View1	$H_{GP_r^1} W_1 (H_{GP_d^1})^T$	$\beta_{(GP_r^1, GP_d^1)}$
View2	$H_{GP_r^2} W_2 (H_{GP_d^2})^T$	$\beta_{(GP_r^2, GP_d^2)}$
View3	$H_{GP_r^3} W_3 (H_{GP_d^3})^T$	$\beta_{(GP_r^3, GP_d^3)}$
View4	$H_{GP_r^4} W_4 (H_{GP_d^4})^T$	$\beta_{(GP_r^4, GP_d^4)}$
View5	$H_{GP_r^5} W_5 (H_{GP_d^5})^T$	$\beta_{(GP_r^5, GP_d^5)}$
View6	$H_{GP_r^6} W_6 (H_{GP_d^6})^T$	$\beta_{(GP_r^6, GP_d^6)}$
View7	$H_{GP_r^7} W_7 (H_{GP_d^7})^T$	$\beta_{(GP_r^7, GP_d^7)}$
View8	$H_{GP_r^8} W_8 (H_{GP_d^8})^T$	$\beta_{(GP_r^8, GP_d^8)}$

known drug-disease associations in our benchmark dataset by  $\{0\%, 10\%, 20\%, 30\%, 40\%, 50\%\}$ . This ensures model stability in sparse networks without relying on unknown associations or external data. The 5-CV results, including *AUC*, *AUPR*, and *F1\_score*, are presented in Table VIII. The Table VIII reveals that MRLHGNN's predictive performance remains consistent even with random removal of known drug-disease associations ( $AUC > 0.94$ ,  $AUPR > 0.59$ ,  $F1\_Score > 0.66$ ). This highlights the robustness of MRLHGNN against variations in drug-disease associations, underscoring its potential for drug repositioning within sparse biological networks.

#### D. Attention Analysis of Views in Automatic Fusion Decoder

To explore the degree of attention for decoding features aggregated from each view-specific graph, we observed the distribution of attention coefficients in the automatic multi-view fusion decoder when training is stable. As shown in Fig. 4, among the sub-views that compose the automatic multi-view fusion decoder, view1 (4 with an attention coefficient greater than  $1/9$ ) and view7 (3 with an attention coefficient greater than  $1/9$ ) receive the most attention under each 5-CV compared to the other sub-views. More specifically, views  $GP_d^1$  and  $GP_d^7$  of diseases and views  $GP_r^1$  and  $GP_r^7$  of drugs contributed more

Attention coefficient for each view in the auto decoder

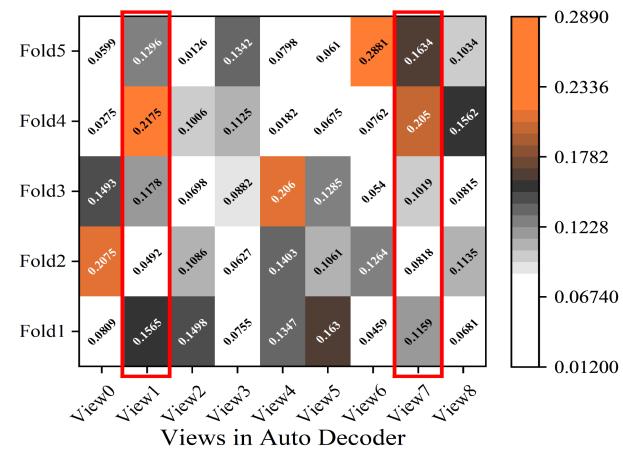


Fig. 4. The distribution of attention coefficient in the MRLHGNN model for decoding from each view in the automatic decoder.

to the MRLHGNN, and in particular, view-specific graph  $GP_d^1$  consisting of the diseases' meta-path  $D \rightarrow D$  and view-specific graph  $GP_r^1$  consisting of the drugs' meta-path  $R \rightarrow R$  produced the greatest impact.

#### E. Ablation Study

To assess the significance and validity of view-specific graphs and sub-module design in MRLHGNN, we conducted relevant ablation experiments by proposing and evaluating four model variants.

- MRLHGNN without view-specific graphs of drugs and diseases (w/o  $GP_r^i$  and  $GP_d^i$ ,  $i \in M$ ): To investigate the effects of view-specific design on the predictive performance of the model, we removed the construction of view-specific graphs for drugs and diseases one by one.
- MRLHGNN without multi-layer feature projection (w/o MLP): Since the multilayer feature projection module is what maintains the feature dimensions of the drug and disease consistent, we designed this ablation experiment using a single-layer feedforward neural network with an output feature dimension of 128 for replacement.

TABLE VI

PERFORMANCE OF TEN METHODS IN TERMS OF *AUC*, *AUPR*, *F1\_score*, *Accuracy*, *Recall*, *Specificity*, AND *Precision* UNDER 5-CV ON DATASET-B

Methods	<i>AUC</i>	<i>AUPR</i>	<i>F1_score</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Specificity</i>	<i>Precision</i>
NTSIM	0.7756 ± 0.0000	0.3802 ± 0.0000	0.4125 ± 0.0000	0.8383 ± 0.0000	0.4957 ± 0.0000	0.8826 ± 0.0000	0.3532 ± 0.0000
BNNR	0.7990 ± 0.0020	0.3960 ± 0.0000	0.4200 ± 0.0000	0.8420 ± 0.0000	0.4980 ± 0.0000	0.8870 ± 0.0000	0.3620 ± 0.0000
NIMCGCN	0.6868 ± 0.0020	0.2710 ± 0.0050	0.3405 ± 0.0069	0.7959 ± 0.0003	0.4601 ± 0.1040	0.8394 ± 0.0003	0.2702 ± 0.0320
HGIMC	0.8078 ± 0.0031	0.4021 ± 0.0010	0.4276 ± 0.0071	0.8463 ± 0.0005	0.5013 ± 0.0086	0.8909 ± 0.0012	0.3727 ± 0.0127
LAGCN	0.8013 ± 0.0078	<b>0.4554 ± 0.0073</b>	<u>0.4493 ± 0.0056</u>	0.8513 ± 0.0044	<u>0.5299 ± 0.0197</u>	0.8928 ± 0.0034	0.3800 ± 0.0137
MiLGNet	0.7824 ± 0.0048	0.3695 ± 0.0053	0.4003 ± 0.0046	0.8294 ± 0.0051	0.4973 ± 0.0082	0.8723 ± 0.0067	0.3351 ± 0.0084
DRWBNCF	0.8005 ± 0.0050	0.3945 ± 0.0082	0.4202 ± 0.0073	0.8424 ± 0.0051	0.4987 ± 0.0037	0.8868 ± 0.0061	0.3631 ± 0.0119
REDDA	0.7402 ± 0.0002	0.4237 ± 0.0010	0.3461 ± 0.0001	0.7546 ± 0.0002	0.5071 ± 0.0060	0.7789 ± 0.0035	0.2490 ± 0.0010
DRHGCN	0.8222 ± 0.0010	0.4057 ± 0.0020	0.3847 ± 0.0028	0.8689 ± 0.0024	0.5085 ± 0.0050	0.9115 ± 0.0033	0.3907 ± 0.0074
MRLHGNN	<b>0.8426 ± 0.0010</b>	<u>0.4310 ± 0.0025</u>	<b>0.4506 ± 0.0031</b>	<b>0.8710 ± 0.0082</b>	<u>0.5741 ± 0.0224</u>	<u>0.9720 ± 0.0118</u>	<b>0.4025 ± 0.0215</b>

<sup>1</sup> Note: The best results are marked in bold and the second-best results are marked as underlined.

(mean ± std).

TABLE VII

PERFORMANCE OF TEN METHODS IN TERMS OF *AUC*, *AUPR*, *F1\_score*, *Accuracy*, *Recall*, *Specificity*, AND *Precision* UNDER 5-CV ON DATASET-C

Methods	<i>AUC</i>	<i>AUPR</i>	<i>F1_score</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Specificity</i>	<i>Precision</i>
NTSIM	0.8236 ± 0.0000	<u>0.4058 ± 0.0000</u>	<b>0.4683 ± 0.0000</b>	0.9830 ± 0.0000	<b>0.5538 ± 0.0000</b>	0.9829 ± 0.0000	<u>0.3056 ± 0.0000</u>
BNNR	0.9421 ± 0.0000	0.3852 ± 0.0000	<u>0.4294 ± 0.0000</u>	0.9810 ± 0.0000	0.4757 ± 0.0000	0.9835 ± 0.0000	0.2913 ± 0.0000
NIMCGCN	0.8331 ± 0.0002	0.0542 ± 0.0001	0.1057 ± 0.0001	0.9630 ± 0.0003	0.2101 ± 0.0001	0.9757 ± 0.0001	0.0706 ± 0.0002
HGIMC	0.9795 ± 0.0003	0.3419 ± 0.0006	0.3643 ± 0.0005	0.9860 ± 0.0001	0.4614 ± 0.0001	0.9906 ± 0.0002	0.3010 ± 0.0000
LAGCN	0.9533 ± 0.0083	<b>0.5497 ± 0.0065</b>	0.2098 ± 0.0019	0.9345 ± 0.0008	0.5072 ± 0.0012	0.9340 ± 0.0009	0.2172 ± 0.0012
MiLGNet	<u>0.9833 ± 0.0013</u>	0.3699 ± 0.0890	0.3756 ± 0.0051	0.9815 ± 0.0003	0.5066 ± 0.0072	<u>0.9909 ± 0.0004</u>	0.2946 ± 0.0072
DRWBNCF	0.7591 ± 0.0024	0.0358 ± 0.0001	0.0672 ± 0.0002	0.9648 ± 0.0019	0.1455 ± 0.0013	0.9720 ± 0.0001	0.0437 ± 0.0020
REDDA	0.9663 ± 0.0019	0.3590 ± 0.0030	0.3684 ± 0.0100	0.9848 ± 0.0351	0.5098 ± 0.0101	0.9889 ± 0.0090	0.2884 ± 0.0210
DRHGCN	0.9563 ± 0.0015	0.2125 ± 0.0050	0.2759 ± 0.0013	0.9845 ± 0.0005	0.3380 ± 0.0002	0.9876 ± 0.0012	0.2331 ± 0.0019
MRLHGNN	<b>0.9842 ± 0.0013</b>	0.3610 ± 0.0007	0.3810 ± 0.0005	<b>0.9863 ± 0.0002</b>	<u>0.5256 ± 0.0012</u>	<b>0.9917 ± 0.0001</b>	<b>0.3135 ± 0.0004</b>

<sup>1</sup> Note: The best results are marked in bold and the second-best results are marked as underlined.

(mean ± std).

TABLE VIII

THE IMPACT OF REMOVING KNOWN DRUG-DISEASE ASSOCIATIONS ON THE PREDICTIVE PERFORMANCE OF MRLHGNN

Removal rate	<i>AUC</i>	<i>AUPR</i>	<i>F1_score</i>
0 %	0.9585 ± 0.0025	0.6645 ± 0.0118	0.7200 ± 0.0004
10 %	0.9531 ± 0.0030	0.6141 ± 0.1122	0.6718 ± 0.0993
20 %	0.9539 ± 0.0020	0.5973 ± 0.0537	0.6669 ± 0.0393
30 %	0.9497 ± 0.0025	0.6784 ± 0.0239	0.7270 ± 0.0174
40 %	0.9477 ± 0.0018	0.6670 ± 0.0040	0.7126 ± 0.0034
50 %	0.9444 ± 0.0002	0.7226 ± 0.0015	0.7582 ± 0.0013

(mean ± std).

- MRLHGNN without Transformer (w/o Transformer): In the MRLHGNN model, the view-specific level node features are not aggregated using the transformer-based sub-module.
- MRLHGNN without attention of auto multi-view fusion decoder (w/o AD\_attention): We removed the view feature decoding attention mechanism from the auto multi-view fusion decoder.

As shown in Fig. 5, each sub-module contributes to the final performance. The Transformer-based view-specific semantic feature fusion module is the most important one. When the MLP module is removed, the model is only affected secondarily by the Transformer-based sub-module. In contrast, removing the attention mechanism in the decoder conversely achieved the least

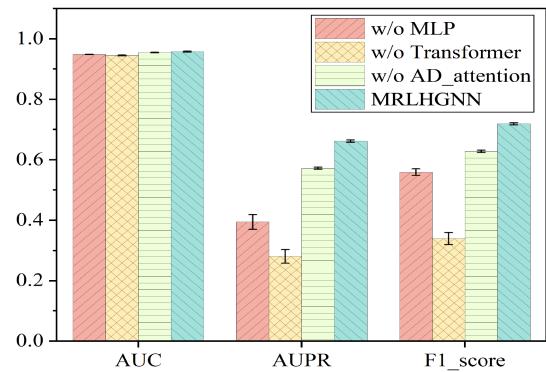


Fig. 5. Results of each ablation method on *AUC*, *AUPR* and *F1\_score* metrics in 5-CV.

impact on the model. Surprisingly, the ablation experimental design showed the flattest effect on *AUC* metrics, but is more significant on *AUPR* and *F1\_score* metrics. According to this ablation study, we can draw the conclusion that a successful drug repositioning prediction model should consider not only feature projection mapping relationships, but also effective feature aggregation mechanisms.

As depicted in Table IX, the MRLHGNN exhibits more pronounced enhancements in *AUPR* and *F1\_score* metrics as the number of view-specific graphs for drugs and diseases increases.

TABLE IX

PERFORMANCE OF ABLATION METHODS IN TERMS OF *AUC*, *AUPR*, *F1\_score*, *Accuracy*, *Recall*, *Specificity*, AND *Precision* UNDER 5-CV ON BENCHMARK DATASET

Methods	<i>AUC</i>	<i>AUPR</i>	<i>F1_score</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Specificity</i>	<i>Precision</i>
w/o $\mathcal{GP}_r^1$ , $\mathcal{GP}_d^1$	0.9546 ± 0.0005	0.5024 ± 0.0269	0.5877 ± 0.0230	0.9932 ± 0.0005	0.7327 ± 0.0249	0.9949 ± 0.0005	0.4910 ± 0.0027
w/o $\mathcal{GP}_r^2$ , $\mathcal{GP}_d^2$	0.9499 ± 0.0073	0.4490 ± 0.0452	0.5304 ± 0.0518	0.9912 ± 0.0021	0.7199 ± 0.0293	0.9934 ± 0.0012	0.4348 ± 0.0342
w/o $\mathcal{GP}_r^3$ , $\mathcal{GP}_d^3$	0.9538 ± 0.0018	0.4721 ± 0.0066	0.5613 ± 0.0582	0.9924 ± 0.0017	0.7240 ± 0.0166	0.9942 ± 0.0016	0.4604 ± 0.0706
w/o $\mathcal{GP}_r^4$ , $\mathcal{GP}_d^4$	0.9549 ± 0.0013	0.5770 ± 0.0251	0.6377 ± 0.0252	0.9943 ± 0.0005	0.7536 ± 0.0273	0.9959 ± 0.0005	0.5532 ± 0.0318
w/o $\mathcal{GP}_r^5$ , $\mathcal{GP}_d^5$	0.9551 ± 0.0008	0.5471 ± 0.0199	0.6180 ± 0.0171	0.9937 ± 0.0005	<u>0.7622 ± 0.0098</u>	0.9953 ± 0.0005	0.5201 ± 0.0256
w/o $\mathcal{GP}_r^6$ , $\mathcal{GP}_d^6$	0.9557 ± 0.0025	0.5850 ± 0.0158	0.6453 ± 0.0095	0.9946 ± 0.0001	0.7339 ± 0.0116	0.9964 ± 0.0005	0.5759 ± 0.0079
w/o $\mathcal{GP}_r^7$ , $\mathcal{GP}_d^7$	0.9548 ± 0.0002	<u>0.5922 ± 0.0024</u>	0.6493 ± 0.0008	0.9947 ± 0.0005	0.7385 ± 0.0058	<u>0.9964 ± 0.0001</u>	<u>0.5793 ± 0.0049</u>
w/o $\mathcal{GP}_r^8$ , $\mathcal{GP}_d^8$	0.9558 ± 0.0017	0.5697 ± 0.0385	0.6389 ± 0.0320	0.9944 ± 0.0006	0.7416 ± 0.0172	0.9961 ± 0.0006	0.5620 ± 0.0414
MRLHGNN	<b>0.9585 ± 0.0025</b>	<b>0.6645 ± 0.0118</b>	<b>0.7200 ± 0.0004</b>	<b>0.9959 ± 0.0030</b>	<b>0.7818 ± 0.0008</b>	<b>0.9974 ± 0.0005</b>	<b>0.6673 ± 0.0010</b>

<sup>1</sup> Note: The best results are marked in bold and the second-best results are marked as underlined.  
(mean ± std).

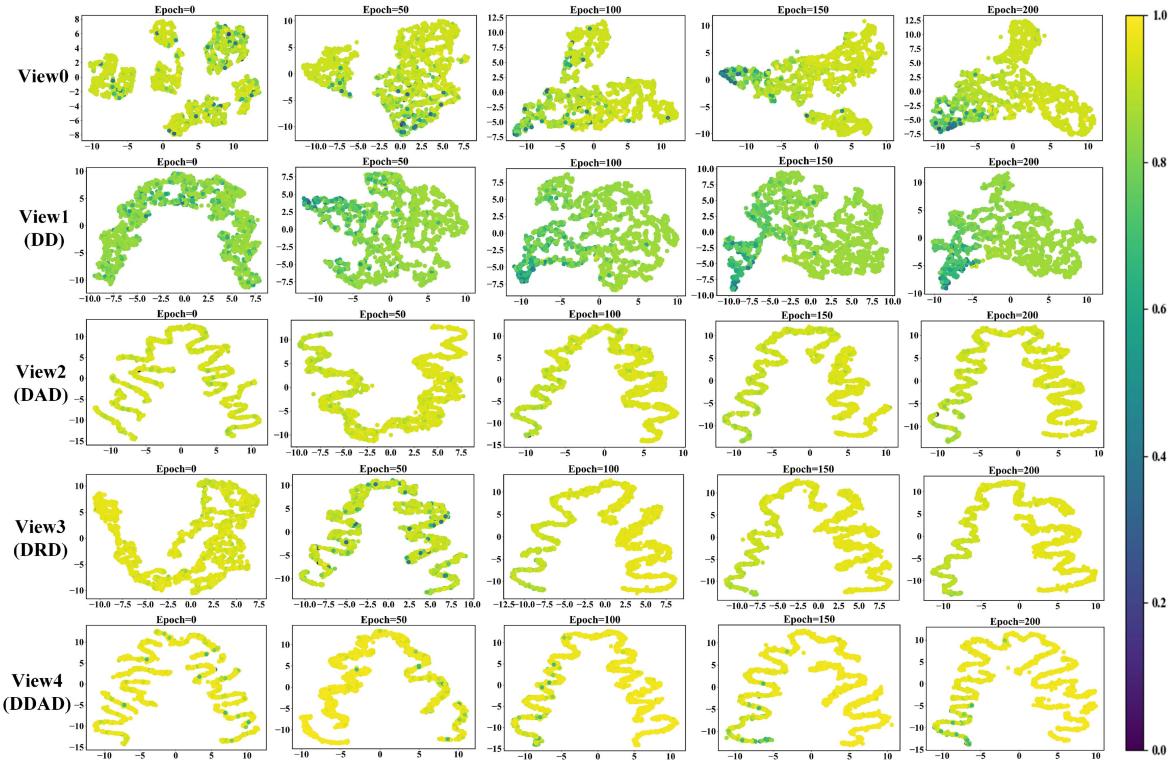


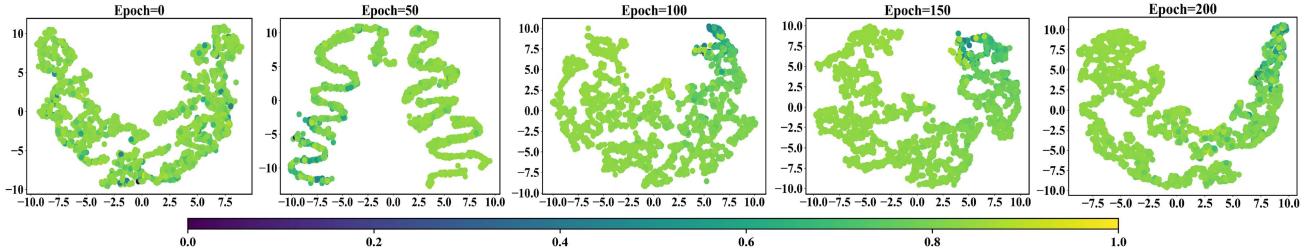
Fig. 6. MDA visualization of training disease features in MRLHGNN.

This clearly demonstrates that combining more meta-path neighbor information into a heterogeneous graph can improve the predictive performance of the model. Furthermore, view-specific graphs consisting of first- and second-hop meta-path neighbor information for drugs and diseases exert a greater influence on the performance of the model. Specifically, without considering the information of second-hop meta-path neighbors and only constructing view-specific graphs from the information of first-hop or higher meta-path neighbors, the *AUPR* and *F1\_score* metrics will reduce by 21.55% and 18.96%, respectively.

#### F. Visualization

To further demonstrate the efficacy of MRLHGNN, we conducted a series of visualization experiments on view-specific

graph layers and transformer feature fusion layers. As shown in Fig. 6, we employed MDA [42] to project the node embeddings onto a two-dimensional feature space of diseases or drugs. Taking diseases as an example, we extracted feature embeddings from different network layers within the disease-specific view of the MRLHGNN model. The visualization experiments reveal that as the number training epochs increases, the color patterns in the feature visualization maps of MRLHGNN exhibit a tendency to cluster and gradual change, while maintaining a continuous and uniform shape. Specifically, in View2's feature visualization, the internal manifold structure of the feature embeddings becomes increasingly distinct with more training epochs, indicating stronger feature learning capability of the model. The comprehensibility arises from the fact that deeper layers of the



**Fig. 7.** Visualizing feature learning of the transformer mechanism in disease view-specific graph using the MDA.

**TABLE X**  
TOP 15 CANDIDATE DRUGS RELATED TO NON-SMALL CELL LUNG CANCER  
PREDICTED BY MRLHGNN

Rank	DrugBank IDs	Drug name	Evidence
1	DB00993	Azathioprine	PMID:23052178
2	DB00563	Methotrexate	PMID:32326159
3	DB01545	Ethyl loflazepate	Unconfirmed
4	DB00704	<u>Naltrexone</u>	PMID:30197847
5	DB00883	Isosorbide dinitrate	PMID:22782331
6	DB04942	Tamibarotene	Unconfirmed
7	DB00618	Demeclacycline	Unconfirmed
8	DB00864	Tacrolimus	PMID:32415507
9	DB00227	Lovastatin	PMID:19760159
10	DB00831	<u>Trifluoperazine</u>	PMID:35625784
11	DB00839	Tolazamide	Unconfirmed
12	DB01021	Trichlormethiazide	PMID:2982333
13	DB00860	Prednisolone	PMID:23761825
14	DB01234	Dexamethasone	PMID:27604504
15	DB00959	Methylprednisolone	PMID:32379397

<sup>1</sup> Note: The predicted drug-disease associations are confirmed by the literature in PubMed and the PMIDs for these are provided.

<sup>2</sup> Note: The underline indicates the examples are described in detail.

view extract higher-level features in order to better accomplish the prediction task.

Additionally, we analyzed the feature learning of the disease view-specific graph using transformer fusion in the model, again using MDA for visualization. As shown in Fig. 7, we observed that with more training epochs, the colors and shapes in the MDA visualization become increasingly regular. This indicates that the network features, after being fused by the transformer mechanism, are orderly distributed in the manifold space. The continuity of colors and shapes in the MDA visualizations suggests that the transformer mechanism maintains the geometric relationships of the feature space well.

## V. CASE STUDIES

To evaluate the actual predictive capability of MRLHGNN, we conducted case studies on non-small cell lung cancer (NSCLC, Mesh ID: D002289) with high morbidity and piroxicam drug (DrugBank ID: DB00554). Before conducting the case study, we considered all known drug-disease associations in the dataset as the training set and the unknown drug-disease associations as the candidate set. Once the model is trained and stable, MRLHGNN obtains the predicted probability of diseases interacting with all drug candidates or drugs interacting with all

**TABLE XI**  
TOP 15 CANDIDATE DISEASES RELATED TO PIROXICAM PREDICTED BY  
MRLHGNN

Rank	MeSH IDs	Disease name	Evidence
1	D001943	Breast neoplasms	PMID:31678245
2	D005879	Tourette syndrome	Unconfirmed
3	D006973	<u>Hypertension</u>	PMID:30349346
4	D008224	Lymphoma, follicular	PMID:12092954
5	D054198	Leukemia-lymphoma	PMID:22285505
6	D015470	Leukemia, myeloid, acute	PMID:27506920
7	D002177	Candidiasis	Unconfirmed
8	D009394	Nephritis, hereditary	PMID:3993703
9	D001249	Asthma	PMID:30742401
10	D003876	<u>Dermatitis</u>	PMID:11674927
11	D001932	Brain neoplasms	Unconfirmed
12	D009298	Nasal Polyps	PMID:12851848
13	D012559	Schizophrenia	Unconfirmed
14	D065631	Rhinitis,allergic	PMID:17975218
15	D005764	Gastroesophageal reflux	PMID:10467627

<sup>1</sup> Note: The predicted drug-disease associations are confirmed by the literature in PubMed and the PMIDs for these are provided.

<sup>2</sup> Note: The underline indicates the examples are described in detail.

diseases candidates. We then rank the drugs'/diseases' candidates based on their predicted probability, where higher-ranked drugs/disease are most likely to treat the diseases or find new indications for drugs.

The results of the top 15 drug candidates predicted by MRLHGNN to have a potential association with NSCLC are shown in Table X. We choose three drugs in Table X to describe them in detail. Azathioprine is a thiopurine, a prodrug that is converted to 6-TG. An outcome of its complex metabolism is the incorporation of 6-TG into DNA by the replication process. Based on the results of a related experiment, Lazarev et al. [43] are enabled to treat patients with NSCLC suffering from ulcerative colitis with azathioprine. Low-dose naltrexone (LDN) can be beneficial as an adjuvant for patients with NSCLC. Miskoff et al. [44] proposed a unique mechanism that allows LDN to enhance the degree of the immune system's ability to affect lung cancer tumors at the cellular level, resulting in longer survival cycles for patients. Trifluoperazine is primarily used to treat schizophrenia. Jeong et al. [45] discovered that a synthetic analogue of Trifluoperazine exhibited a strong activity against lung cancer in clinical treatment.

The top15 disease candidates for piroxicam are shown in Table XI. We also selected three disease candidates for detailed

description. Piroxicam is a non-steroidal anti-inflammatory drug, Thabet et al. [46] determined that piroxicam inhibits inflammation-driven breast neoplasms from RAS and PAR-4 signaling. Besides, Mazzilli et al. [47] found that topical 0.8% piroxicam and 50+ sunscreen filters reduced the location of lesions in actinic keratoses for hypertensive patients. In addition, through actual cases, Trujillo et al. [48] found that patients with dermatitis taking piroxicam should avoid using it with other medications.

Overall, 11 of the top 15 drugs'/diseases' candidates for prediction have been proven in the literature, which can further illustrate the reliable performance of MRLHGNN for drug repositioning.

## VI. CONCLUSION

Drug repositioning can effectively improve the efficiency of drug development and disease treatment. It is interdependent with drug combination therapy in modern medicine and provides a new way for disease treatment [49], [50]. In this paper, we propose a drug repositioning method (MRLHGNN) with a transformer-based view-specific graph-level feature aggregation mechanism and a multi-view automatic fusion decoder. In heterogeneous bioinformatics networks, we first construct view-specific graphs based on meta-paths. The definition of meta-path types is based on a pre-exploration of the structure and domain expertise of bio-heterogeneous networks, making our model interpretable. Then, we obtain the node-structure feature vector representation at meta-path level and the low-dimensional semantic vector representation of graph-level features for predicting drug-disease associations using view-specific graph-weighted aggregation mechanism and transformer mechanism. The transformer mechanism captures the interdependencies between view-specific graphs. To some extent, the multi-view automatic fusion decoder reflects the impact of encoder views on the prediction performance of drug-disease associations. To validate the effectiveness of MRLHGNN, we compared it with 9 state-of-the-art prediction methods for drug repositioning on a benchmark dataset. The results indicate that MRLHGNN exhibits significant performance enhancement and competitive. Moreover, case studies have illustrated that MRLHGNN can be accepted as a reliable tool for drug repositioning.

Although MRLHGNN achieves SOTA performance in the comparison method, there are still some problems that deserve deeper study, such as the imbalance between the number of positive and negative samples, and the selection strategy of reliable negative samples. In the future, we will continue to address these shortcomings through a comprehensive exploration of supervised contrast learning [51] and hypergraph learning [52].

## REFERENCES

- [1] H. S. Chan, H. Shan, T. Dahoun, H. Vogel, and S. Yuan, "Advancing drug discovery via artificial intelligence," *Trends Pharmacological Sci.*, vol. 40, no. 8, pp. 592–604, 2019.
- [2] H. Liu, Y. Zhao, L. Zhang, and X. Chen, "Anti-cancer drug response prediction using neighbor-based collaborative filtering with global effect removal," *Mol. Ther.-Nucleic Acids*, vol. 13, pp. 303–311, 2018.
- [3] N.-N. Guan, Y. Zhao, C.-C. Wang, J.-Q. Li, X. Chen, and X. Piao, "Anti-cancer drug response prediction in cell lines using weighted graph regularized matrix factorization," *Mol. Ther.-Nucleic Acids*, vol. 17, pp. 164–174, 2019.
- [4] T.-H. Li, C.-C. Wang, L. Zhang, and X. Chen, "SNRMPACDC: Computational model focused on siamese network and random matrix projection for anticancer synergistic drug combination prediction," *Brief. Bioinf.*, vol. 24, no. 1, 2023, Art. no. bbac503.
- [5] X. Chen and L. Huang, "Computational model for disease research," *Brief. Bioinf.*, vol. 24, no. 1, 2023, Art. no. bbac615.
- [6] X. Chen, C. Zhou, C.-C. Wang, and Y. Zhao, "Predicting potential small molecule–miRNA associations based on bounded nuclear norm regularization," *Brief. Bioinf.*, vol. 22, no. 6, 2021, Art. no. bbab328.
- [7] L. Peng, Y. Tu, L. Huang, Y. Li, X. Fu, and X. Chen, "DAESTB: Inferring associations of small molecule–miRNA via a scalable tree boosting model based on deep autoencoder," *Brief. Bioinf.*, vol. 23, no. 6, 2022, Art. no. bbac478.
- [8] H. Wang, F. Huang, Z. Xiong, and W. Zhang, "A heterogeneous network-based method with attentive meta-path extraction for predicting drug-target interactions," *Brief. Bioinf.*, vol. 23, no. 4, 2022, Art. no. bbac184.
- [9] L. Zhang, C.-C. Wang, and X. Chen, "Predicting drug–target binding affinity through molecule representation block based on multi-head attention and skip connection," *Brief. Bioinf.*, vol. 23, no. 6, 2022, Art. no. bbac468.
- [10] W. Lan et al., "IGNSCDA: Predicting CircRNA-disease associations based on improved graph convolutional network and negative sampling," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 6, pp. 3530–3538, Nov./Dec. 2021.
- [11] L. Wang, L. Wong, Z.-H. You, D.-S. Huang, X.-R. Su, and B.-W. Zhao, "NSECDA: Natural semantic enhancement for circRNA-disease association prediction," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 10, pp. 5075–5084, Oct. 2022.
- [12] W. Liu, T. Tang, X. Lu, X. Fu, Y. Yang, and L. Peng, "MPCLCDA: Predicting circRNA-disease associations by using automatically selected meta-path and contrastive learning," *Brief. Bioinf.*, vol. 24, 2023, Art. no. bbad227.
- [13] W. Zhang et al., "Predicting drug-disease associations based on the known association bipartite network," in *2017 IEEE Int. Conf. Bioinf. Biomed.*, 2017, pp. 503–509.
- [14] M. Yang, H. Luo, Y. Li, and J. Wang, "Drug repositioning based on bounded nuclear norm regularization," *Bioinformatics*, vol. 35, no. 14, pp. i455–i463, 2019.
- [15] M. Yang, L. Huang, Y. Xu, C. Lu, and J. Wang, "Heterogeneous graph inference with matrix completion for computational drug repositioning," *Bioinformatics*, vol. 36, no. 22/23, pp. 5456–5464, 2020.
- [16] X. Ji, J. M. Freudenberg, and P. Agarwal, "Integrating biological networks for drug target prediction and prioritization," in *Computational Methods for Drug Repurposing*. Berlin, Germany: Springer, 2019, pp. 203–218.
- [17] Z. Yu, F. Huang, X. Zhao, W. Xiao, and W. Zhang, "Predicting drug-disease associations through layer attention graph convolutional network," *Brief. Bioinf.*, vol. 22, no. 4, 2021, Art. no. bbaa243.
- [18] L. Cai et al., "Drug repositioning based on the heterogeneous information fusion graph convolutional network," *Brief. Bioinf.*, vol. 22, no. 6, 2021, Art. no. bbab319.
- [19] Y. Meng, C. Lu, M. Jin, J. Xu, X. Zeng, and J. Yang, "A weighted bilinear neural collaborative filtering approach for drug repositioning," *Brief. Bioinf.*, vol. 23, no. 2, 2022, Art. no. bbab581.
- [20] Y. Gu, S. Zheng, Q. Yin, R. Jiang, and J. Li, "REDDA: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction," *Comput. Biol. Med.*, vol. 150, 2022, Art. no. 106127.
- [21] Y. Gu, S. Zheng, B. Zhang, H. Kang, and J. Li, "Milgnet: A multi-instance learning-based heterogeneous graph network for drug repositioning," in *2022 IEEE Int. Conf. Bioinf. Biomed.*, 2022, pp. 430–437.
- [22] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.
- [23] Z. Shao, Y. Xu, W. Wei, F. Wang, Z. Zhang, and F. Zhu, "Heterogeneous graph neural network with multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11476–11488, Nov. 2023.
- [24] H. Fu, F. Huang, X. Liu, Y. Qiu, and W. Zhang, "MVGNC: Data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks," *Bioinformatics*, vol. 38, no. 2, pp. 426–434, 2022.

- [25] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "PREDICT: A method for inferring novel drug indications with application to personalized medicine," *Mol. Syst. Biol.*, vol. 7, no. 1, 2011, Art. no. 496.
- [26] H. Luo et al., "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [27] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.
- [28] A. P. Davis et al., "Comparative toxicogenomics database (CTD): Update 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D1138–D1143, 2021.
- [29] D. S. Wishart et al., "DrugBank: A knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Res.*, vol. 36, no. suppl\_1, pp. D901–D906, 2008.
- [30] C. v. Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel, "STRING: A database of predicted functional associations between proteins," *Nucleic Acids Res.*, vol. 31, no. 1, pp. 258–261, 2003.
- [31] U. Consortium et al., "UniProt: The universal protein knowledgebase," *Nucleic Acids Res.*, vol. 46, no. 5, pp. D115–119, 2018.
- [32] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010.
- [33] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of go terms," *Bioinformatics*, vol. 23, no. 10, pp. 1274–1281, 2007.
- [34] Z.-H. Guo, Z.-H. You, D.-S. Huang, H.-C. Yi, Z.-H. Chen, and Y.-B. Wang, "A learning based framework for diverse biomolecule relationship prediction in molecular association network," *Commun. Biol.*, vol. 3, no. 1, 2020, Art. no. 118.
- [35] M. Y. Wang, "Deep graph library: Towards efficient and scalable deep learning on graphs," in *Proc. ICLR Workshop Representation Learn. Graphs Manifolds*, 2019, pp. 1–7. [Online]. Available: <https://par.nsf.gov/biblio/10311680>
- [36] Y. Sun, D. Zhu, H. Du, and Z. Tian, "MHNF: Multi-hop heterogeneous neighborhood information fusion graph representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7192–7205, Jul. 2023.
- [37] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–15, vol. 30.
- [38] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *Proc. ICLR Workshop Representation Learn.*, 2014, pp. 1–12.
- [39] J. Li, S. Zhang, T. Liu, C. Ning, Z. Zhang, and W. Zhou, "Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction," *Bioinformatics*, vol. 36, no. 8, pp. 2538–2546, 2020.
- [40] L. Peng, C. Yang, L. Huang, X. Chen, X. Fu, and W. Liu, "RNMFPLP: Predicting circRNA–disease associations based on robust nonnegative matrix factorization and label propagation," *Brief. Bioinf.*, vol. 23, no. 5, 2022, Art. no. bbac155, doi: [10.1093/bib/bbac155](https://doi.org/10.1093/bib/bbac155).
- [41] L. Peng, C. Yang, Y. Chen, and W. Liu, "Predicting CircRNA-disease associations via feature convolution learning with heterogeneous graph attention network," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 6, pp. 3072–3082, Jun. 2023.
- [42] M. T. Islam et al., "Revealing hidden patterns in deep neural network feature space continuum via manifold learning," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 8506.
- [43] I. Lazarev, N. Sion-Vardy, and S. Ariad, "EML4-ALK-positive non-small cell lung cancer in a patient treated with azathioprine for ulcerative colitis," *Tumori J.*, vol. 98, no. 4, pp. e98–e101, 2012.
- [44] J. A. Miskoff and M. Chaudhri, "Low dose naltrexone and lung cancer: A case report and discussion," *Cureus*, vol. 10, no. 7, 2018, Art. no. e2924.
- [45] J. Y. Jeong et al., "Trifluoperazine and its analog suppressed the tumorigenicity of non-small cell lung cancer cell; applicability of antipsychotic drugs to lung cancer treatment," *Biomedicines*, vol. 10, no. 5, 2022, Art. no. 1046.
- [46] N. A. Thabet, N. El-Guendy, M. M. Mohamed, and S. A. Shouman, "Suppression of macrophages-induced inflammation via targeting ras and par-4 signaling in breast cancer cell lines," *Toxicol. Appl. Pharmacol.*, vol. 385, 2019, Art. no. 114773.
- [47] S. Mazzilli et al., "Effects of topical 0.8% piroxicam and 50 sunscreen filters on actinic keratoses in hypertensive patients treated with or without photosensitizing diuretic drugs: An observational cohort study," *Clin. Cosmetic Investigational Dermatol.*, vol. 11, pp. 485–490, 2018.
- [48] M. Trujillo et al., "Piroxicam-induced photodermatitis. Cross-reactivity among oxicams. A case report," *Allergologia et immunopathologia*, vol. 29, no. 4, pp. 133–136, 2001.
- [49] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: Predicting synergistic drug combinations based on semi-supervised learning," *PLoS Comput. Biol.*, vol. 12, no. 7, 2016, Art. no. e1004975.
- [50] X. Chen et al., "Drug–target interaction prediction: Databases, web servers and computational models," *Brief. Bioinf.*, vol. 17, no. 4, pp. 696–712, 2016.
- [51] Z. Xiong et al., "Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, pp. 5339–5347.
- [52] Q. Ning et al., "AMHMDA: Attention aware multi-view similarity networks and hypergraph learning for miRNA–disease associations identification," *Brief. Bioinf.*, vol. 24, no. 2, 2023, Art. no. bbad094.