

Predicting Anti-Cancer Drug Response Based on Hypergraph Representation Learning

Wei Peng*, Xinyue Xu, Jiangzhen Lin, Gong Chen, Wei Dai*, Xiaodong Fu, Li Liu, Lijun Liu, Ning Yu

Abstract—Accurate prediction of drug responses is critical for advancing personalized cancer therapies. Although current graph neural network (GNN)-based approaches predominantly focus on pairwise interactions between cell lines and drugs, they often neglect the potential of higher-order interactions. In this study, we present HRLCDR, a novel computational framework that utilizes Hypergraph Representation Learning to predict Cancer Drug Responses. HRLCDR begins by constructing hypergraphs for both cell lines and drugs and then processes through low-pass and high-pass hypergraph convolutions, allowing the model to extract both common and different features from the complex higher-order interactions between cell lines and drugs. After that, HRLCDR constructs a heterogeneous graph using known cell line responses to drugs. Parallel heterogeneous graph convolution operations are then employed to extract primary interaction features between cell lines and drugs from these associations. Finally, HRLCDR integrates the features learned from both the hypergraphs and the heterogeneous graph, predicting drug response via Classifiers. We evaluated HRLCDR’s performance on two major cancer drug response datasets: the Cancer Drug Sensitivity Data (GDSC) and the Cancer Cell Line Encyclopedia (CCLE). The results demonstrate that HRLCDR outperforms current state-of-the-art methods, underscoring its potential to enhance the accuracy and reliability of cancer drug response predictions.

Index Terms—Hypergraph Representation Learning, Hypergraph Convolution, Drug sensitivity prediction.

I. INTRODUCTION

ACCURATELY predicting anti-cancer drug response plays a critical role in precision medicine, enabling the customization of cancer therapies based on the unique genetic profiles of individual patients. Traditional approaches typically rely on labor-intensive experiments and protracted clinical trials to assess drug efficacy [1]. The emergence of high-throughput sequencing technologies, alongside initiatives

W. Peng and W. Dai are with Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, P. R. China and Computer Technology Application Key Lab of Yunnan Province, Kunming University of Science and Technology, Kunming, Yunnan 650500, P. R. China. E-mail: weipeng1980@gmail.com and daiwei@kust.edu.cn.

X.Y. Xu, J.Z. Lin and G. Chen are with Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, P. R. China.

X.D. Fu, L. Liu and L.J. Liu are with Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, P. R. China.

N. Yu is with the State University of New York, The College at Brockport, Department of Computing Sciences, 350 New Campus Drive, Brockport, NY 14422.

Code available: <https://github.com/weiba/HRLCDR>

*Corresponding author

like the Cancer Cell Line Encyclopedia (CCLE) [2] and the Genomics of Drug Sensitivity in Cancer (GDSC) [3], has significantly expanded the availability of drug response data for cancer research. These resources, when combined with advanced computational models, allow for more precise predictions of therapeutic outcomes. By utilizing such models, researchers can optimize drug regimens, enhance treatment efficacy, expedite clinical translation, and mitigate both financial costs and adverse effects. Collectively, these advancements promise improved patient survival rates and a more personalized, data-driven approach to cancer therapy. Currently, numerous advanced computational models predict anti-cancer drug responses by integrating cell line and drug features. These models can be categorized into two groups based on their feature fusion strategies: machine learning-based methods and network embedding-based methods.

Machine learning-based approaches typically transform and fuse cell line and drug features, utilizing classifiers such as Support Vector Machines (SVM), Multilayer Perceptron (MLP) [4], Deep Forest [5], and Convolutional Neural Networks (CNN) [6] to predict drug responses. In these methods, MLPs, autoencoders (AE), and Cascade Forest are frequently employed to extract cell line features from multi-omics datasets, such as gene expression profiles, DNA methylation data, and genomic mutation data. Concurrently, drug features are derived from molecular fingerprints using MLPs or from drug molecular graphs using Graph Neural Networks (GNNS) [6]. These features are subsequently combined and input into the model’s classifier to predict cell line-drug interactions and drug responses. However, a significant limitation of these approaches is their inability to fully capture the complex relationships between drugs and cell lines, which are essential for accurate predictions.

Network embedding-based methods operate under the assumption that similar cell lines exhibit analogous responses to similar drugs. These approaches typically construct heterogeneous networks, where cell lines and drugs are represented as nodes and known responses are represented as edges. The primary objective is to predict drug responses for novel cell lines by leveraging both the topological and attribute information embedded within the network. These models learn low-dimensional embeddings for cell lines and drugs in a latent space while preserving the network structure and associated node attributes. Ultimately, they infer unknown cell line-drug responses by reconstructing the relationships between drugs and cell lines using these latent embeddings. Graph Neural

Networks (GNNs) have emerged as a powerful deep learning tool for processing graphs, particularly in the context of cancer drug response prediction. GNNs effectively integrate node attributes with graph structural information, facilitating enriched feature learning for both drugs and cell lines. A notable example is MOFGCN, proposed by Peng et al. [7], which predicts drug responses in cell lines by leveraging multi-omics fusion and graph convolutional networks. MOFGCN constructs a heterogeneous network by integrating cell line similarity, drug similarity, and known cell line-drug associations. The model then employs graph convolutional operations to aggregate features from neighboring nodes, enabling the learning of latent features for cancer cell lines and drugs. Finally, these features predict drug responses by reconstructing the cell line-drug correlation matrix. GraphCDR [8] enhances drug response prediction by constructing two distinct network views about whether cell line is sensitive or resistant to specific drugs. It simultaneously integrates contrastive learning within the graph convolutional module to improve the quality of learned node features, resulting in more accurate predictions. Similarly, NIHGCN [9] employs two separate graph convolutions to independently learn features for cell lines and drugs, and incorporates element-wise interactions between nodes and their neighbors to refine feature representation and prediction accuracy. In addition to the similarity of different types of nodes in the heterogeneous network, TSGCNN [10] considers the similarity between homogeneous nodes. It utilizes a dual-space graph convolutional network to learn features for cell lines and drugs from their respective heterogeneous and homogeneous neighbors. The model employs a linear correlation coefficient decoder alongside the final representations to reconstruct the cell line-drug correlation matrix for drug response prediction. However, network embedding methods typically rely on direct connections between nodes in both heterogeneous and homogeneous networks, limiting their ability to capture higher-order interactions. To overcome this limitation, it is essential to develop methods that can effectively learn features for cell lines and drugs from higher-order neighbors. Propagating cell response information along paths within these networks enables a more comprehensive representation of node interactions. By incorporating both direct and higher-order connections, it can capture more complex relationships and dependencies, which are crucial for understanding the intricate dynamics between cell lines, drugs, and their responses. Additionally, integrating both homogeneous and heterogeneous data, to enhance the understanding of underlying biological processes.

Hypergraphs extend traditional graph models by allowing edges to connect an arbitrary number of vertices sharing similar attributes. Unlike conventional graphs, which only represent direct relationships between pairs of nodes, hypergraphs effectively model higher-order interactions. Hypergraphs can capture intricate, multi-faceted dependencies in biological systems and represent high-order correlations and complex interactions [11], [12], [13]. To address these challenges, we propose a novel computational framework, HRLCDR, which leverages hypergraph representation learning to predict cell line responses to drugs. In this approach, hypergraphs are con-

structed between cell lines and drugs based on their similarities to capture both commonalities and distinctions in the features of cell lines and drugs. The framework further employs a combination of low-pass and high-pass hypergraph convolutions to extract shared and distinct features from high-order interactions among the hypergraph's nodes. HRLCDR subsequently constructs a heterogeneous graph based on existing cell line-drug response data. It employs parallel heterogeneous graph convolutional layers to extract distinct features for both cell lines and drugs from the associations within the heterogeneous graph. Finally, the model reconstructs cell line-drug associative relationships by integrating features learned from both the hypergraph and the heterogeneous graph, thereby improving the accuracy of drug response predictions. Experimental results validate the effectiveness of HRLCDR in predicting anticancer drug responses and demonstrating strong performance on the GDSC and CCLE datasets.

In summary, our model provides the following contributions:

- 1) We propose HRLCDR, a computational method for predicting cell line responses to drugs using hypergraph representation learning. The method constructs drug and cell line hypergraph based on the similarity matrices of drugs and cell lines and their interactions. For drug and cell line features, low-pass and high-pass hypergraph convolutions are applied to capture common and different features among nodes.
- 2) We constructs a heterogeneous graph network where cell lines and drugs are represented as nodes. Then, we apply parallel heterogeneous graph convolutions to extract node features. Furthermore, the model predicts cell line response strengths to drugs by combining linear correlation coefficients and multi-feature fusion.
- 3) Extensive experimental results demonstrate that the HRLCDR algorithm outperforms state-of-the-art algorithms across various metrics, highlighting its effectiveness in predicting anticancer drug responses.

II. MATERIALS AND METHODS

A. Materials

In this study, we analyzed drug response data obtained from two prominent sources: the Genomics of Drug Sensitivity in Cancer (GDSC) [3] and the Cancer Cell Line Encyclopedia (CCLE) [2]. From the GDSC database, we utilized Tables S4A and S5C, which provided detailed information on drug sensitivity and resistance. Following established protocols [9], a drug was classified as "sensitive" for a specific cell line if its IC₅₀ value in Table S4A exceeded the threshold defined in Table S5C. Conversely, drugs with IC₅₀ values below this threshold were deemed "resistant." The preprocessed GDSC dataset contained drug response data for 962 cancer cell lines treated with 265 drugs, comprising 20,851 sensitive and 156,512 resistant samples.

For the CCLE dataset, we constructed an experimental benchmark using data from 11,670 drug assays across various cell lines. To ensure comparability, IC₅₀ values were

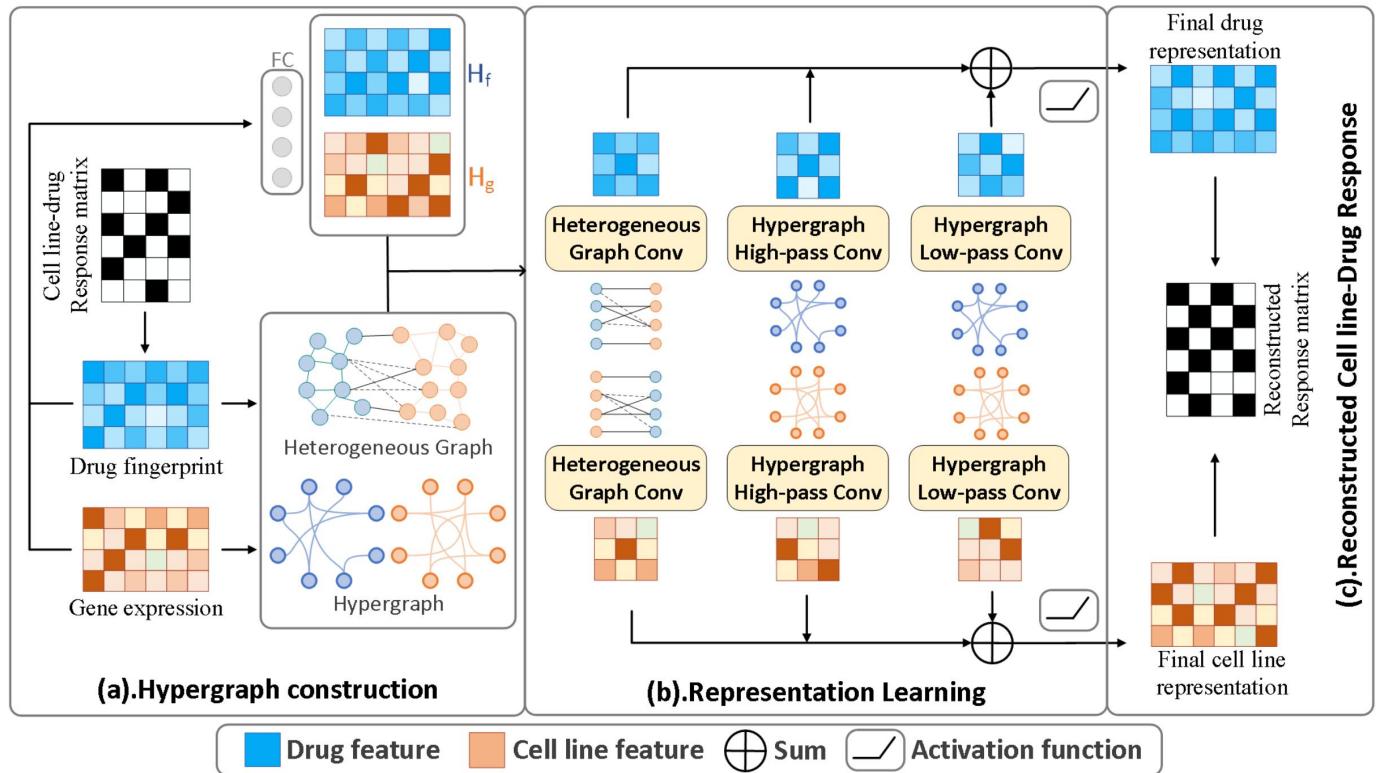


Fig. 1. The framework of HRLCDR.

normalized using established methodologies [10]. Drug sensitivity was defined as a normalized IC₅₀ value below -0.8, while values above this threshold indicated resistance. The preprocessed CCLE dataset contained drug response data for 24 drugs across 436 cell lines, including 1,696 sensitive and 8,768 resistant samples.

B. Methods Overview

HRLCDR predicts cell line-drug responses through a three-step process (Figure 1). Firstly, separate hypergraphs are constructed for cell lines and drugs, where the edges represent the similarity relationships between these entities. Subsequently, HRLCDR applies both low-pass and high-pass hypergraph convolutions to these hypergraphs. By simultaneously considering the similarity (low-pass) and dissimilarity (high-pass) between neighboring nodes, it captures the higher-order interactions between the nodes. Simultaneously, parallel heterogeneous graph convolutions are employed to learn features from the known associations between cell lines and drugs, further refining the representation of both cell line and drug characteristics. Ultimately, HRLCDR integrates the feature representations obtained from both the hypergraph convolutions (low-pass and high-pass) and the heterogeneous graph convolution. These integrated features are used to calculate the linear correlation coefficients between cell lines and drugs, predicting the strength of a cell line's response to a drug.

C. Hypergraph Construction

We propose a novel approach to constructing drug and cell line hypergraphs based on the premise that similar drugs tend

to exhibit similar responses in similar cell lines. This approach utilizes similarity matrices for drugs and cell lines, along with a cell line-drug response matrix, to capture higher-order associations between drugs and cell lines [14]. The method not only considers pairwise interactions but also integrates indirect higher-order relationships.

The first step involves calculating the drug similarity matrix S_d using the Jaccard similarity coefficient. The Jaccard similarity coefficient quantifies the degree of overlap between these fingerprint sets, reflecting the chemical similarity between the two drugs. For each pair of drugs, the similarity of their drug fingerprints is computed, as shown in Equation (1):

$$S_{d(ij)} = \begin{cases} Sim_{d(ij)} & j \in K_i, \\ 0 & j \notin K_i \end{cases} \quad (1)$$

Where, $Sim_{d(ij)} = \frac{|d_i \cap d_j|}{|d_i \cup d_j|}$ denotes the Jaccard similarity between drug fingerprints, K_i represents the set of the top K drugs most similar to drug i , with K set to 10.

Afterwards, the cell line similarity matrix S_c is computed using gene expression profiles. The similarity between cell lines is quantified using an exponential similarity function, as defined in Equation (2):

$$S_{c(ij)} = \begin{cases} Sim_{c(ij)} & j \in T_i, \\ 0 & j \notin T_i \end{cases} \quad (2)$$

Here, $Sim_{c(ij)} = e^{-\frac{\|X_i - X_j\|_2}{2\varepsilon^2}}$ represents the exponential similarity of gene expression between cell lines, where X_i denotes the normalized gene expression values of the cell

lines i , and T_i denotes the set of the top T cell lines most similar to the cell line i , with T set to 10. By choosing different similarity calculation methods by the nature of the data, the Jaccard similarity is used for discrete data such as drug fingerprints, while the exponential similarity is applied to continuous data like gene expression profiles of cell lines, thereby improving the accuracy of drug response predictions by capturing underlying similarities.

Let A denote the existing cell line-drug response matrix. Drawing upon the methodology presented in [13], we construct the hypergraph for drug H_d using the following Equations to capture higher-order associations among drugs. This approach facilitates the integration of complex, multi-dimensional relationships between drugs, thereby improving the predictive accuracy of drug responses by accounting for both direct pairwise interactions and indirect associations.

$$H_{d1} = (S_d \cdot S_d) \odot S_d \quad (3)$$

$$H_{d2} = (A^T \cdot A) \odot S_d \quad (4)$$

$$H_{d3} = (A^T \cdot S_c \cdot A) \odot S_d + ((A^T \cdot S_c \cdot A) \odot S_d)^T \quad (5)$$

$$H_d = H_{d1} + H_{d2} + H_{d3} \quad (6)$$

Where, H_{d1} represents the drug similarity matrix, H_{d2} reflects the similarity of cell line responses associated with these drugs, H_{d3} captures indirect cell line correlations between similar drugs. By integrating these higher-order matrices, the drug hypergraph H_d is constructed. Correspondingly, we construct the cell line hypergraph H_c to model higher-order relationships between cell lines using the following equations:

$$H_{c1} = (S_c \cdot S_c) \odot S_c \quad (7)$$

$$H_{c2} = (A \cdot A^T) \odot S_c \quad (8)$$

$$H_{c3} = (A \cdot S_d \cdot A^T) \odot S_c + ((A \cdot S_d \cdot A^T) \odot S_c)^T \quad (9)$$

$$H_{c3} = (S_c \cdot A^T) \cdot (S_c \cdot A^T)^T \quad (10)$$

$$H_c = H_{c1} + H_{c2} + H_{c3} + H_{c4} \quad (11)$$

Where, H_{c1} stores the cell line similarities between similar cell lines, H_{c2} represents similar drug responses between similar cell lines, H_{c3} indicates an indirect similar cell line correlation between similar cell lines, while H_{c4} completes the direct correlation relationships between the cell lines. Ultimately, a summation of these higher-order matrices yields the cell line hypergraph, denoted as H_c .

D. Initial Features of Cell Lines/Drugs

Primarily, we normalize the expression levels of each gene across all cell lines using Equation (12), to obtain \overline{expr}_i :

$$\overline{expr}_i = \frac{(expr_i - u_i)}{\sigma_i} \quad (12)$$

Here, $expr_i$ represents the expression values of the i th gene across all cell lines, with a mean of u_i and a standard deviation of σ_i . Consequently, we derive the initial cell line features $X_g \in \mathbb{R}^{m \times d_g}$ through Equation (13):

$$X_g = C_g \cdot W_g \quad (13)$$

Where, $C_g \in \mathbb{R}^{m \times d_{cg}}$ represents the normalized gene expression values of cell lines, which are obtained via Equation (12). m represents the count of cell lines, d_{cg} refers to the number of genes, and $W_g \in \mathbb{R}^{d_{cg} \times d_g}$ denotes the set of learnable parameters. Furthermore, d_g signifies the embedding dimension of the transformed cell line features.

To derive the initial drug features $X_f \in \mathbb{R}^{n \times d_f}$, a linear transformation is applied to the drug molecular fingerprint matrix, as described in Equation (14):

$$X_f = D_f \cdot W_f \quad (14)$$

The drug molecular fingerprint matrix is denoted as $D_f \in \mathbb{R}^{n \times d_d}$, where n indicates the total number of drugs, d_d denotes the length of the drug molecular fingerprint features. The transformation relies on a set of learnable parameter for the transformation of drug features, represented as $W_f \in \mathbb{R}^{d_d \times d_f}$, which project the fingerprint features into an embedding space with a dimensionality of d_f .

E. Low-pass and High-pass Hypergraph Convolutional Representation Learning

Hypergraphs are capable of capturing higher-order correlations among nodes, facilitating more nuanced representations of interactions. To identify both common and different features among the nodes, we apply low-pass and high-pass filters to hypergraphs. In Graph Neural Networks (GNNs), high-pass and low-pass filters are employed to process graph signals and differentiate between frequency components, and these concepts are derived from signal processing. Specifically, a high-pass filter permits high-frequency signals while blocking low-frequency ones. In the background of graphs, low-frequency signals typically represent smooth or slowly varying features, indicating that neighboring nodes exhibit similar characteristics. Conversely, high-frequency signals are associated with abrupt changes or anomalies in the graph, suggesting significant disparities between a node's features and those of its neighbors. By applying both low-pass and high-pass filters to hypergraphs, it is possible to effectively extract both common and different features among nodes.

Given a cell line hypergraph H_c , its normalized graph Laplacian operator L_c is defined as $L_c = I_N - D^{-1}H_cD^{-1}$, where $D = \sum_j H_{c(i,j)} + 1$. L_c summarizes the signal differences between neighboring nodes, functioning as a high-pass filter that highlights variations across the hypergraph H_c [15]. Consequently, we employ high-pass graph convolution operations, as outlined in Equations (15) and (16), to extract relevant cell line features $X_{ch}^{(L_h)}$. This method facilitates the identification of significant variations within the hypergraph structure, enhancing the model's ability to predict drug responses by emphasizing the most informative signal components.

$$Z_{ch}^{(L_h)} = L_c X_{ch}^{(L_h-1)} W_{ch}^{(L_h-1)} \quad (15)$$

$$X_{ch}^{(L_h)} = \alpha Z_{ch}^{(L_h)} + \beta (Z_{ch}^{(L_h)} \odot X_c^{(0)}) \quad (16)$$

Let $X_c^{(0)} = X_g$ represent the initial feature of cell line, as defined in Equation (13). To mitigate the risk of over-smoothing, we incorporate the dot product between the initial node features and the features obtained after applying the convolution operation. Additionally, W_h represents the learnable parameters associated with the high-pass convolution. α and β are hyperparameters that control the relative contributions of feature aggregation and residual connections, respectively. Finally, L_h denotes the number of high-pass Hypergraph convolutional layers in the network. This formulation ensures an appropriate balance between the learned features and the original input, thereby enhancing the stability and predictive accuracy of the model in cancer drug response prediction.

The low-pass convolution operation $S_c = D^{-\frac{1}{2}} H_c D^{-\frac{1}{2}}$ computes the average of neighboring node features, serving as a low-pass filter that smooths the hypergraph structure H_c [16]. This approach facilitates the aggregation of local feature information, allowing the model to capture smooth variations within the graph. We obtain the node features of the low-pass convolution $X_{cl}^{(L_l)}$ through Equation (17) and (18):

$$Z_{cl}^{(L_l)} = S_c X_{cl}^{(L_l-1)} W_{cl}^{(L_l-1)} \quad (17)$$

$$X_{cl}^{(L_l)} = \alpha Z_{cl}^{(L_l)} + \beta (Z_{cl}^{(L_l)} \odot X_c^{(0)}) \quad (18)$$

Where W_{cl} represents the learnable parameters for low-pass convolution and L_l denotes the number of low-pass Hypergraph convolutional layers in the network. Similarly, using Equations (16) and (18), we can compute the hypergraph high-pass features $X_{dh}^{(L_h)}$ and low-pass features $X_{dl}^{(L_l)}$ for the drugs within the drug hypergraph. To achieve this, we replace L_c and S_c with L_d and S_d that are computed from the drug hypergraph H_d . Additionally, $X_d^{(0)} = X_f$ denotes the initial drug features obtained from Equation (14). This approach ensures the effective application of both high-pass and low-pass filters to the drug-specific data, facilitating the extraction of relevant features for cancer drug response prediction.

F. Heterogeneous Graph Convolutional Representation Learning

Heterogeneous network was constructed to model associations between cell lines and drugs, where nodes represent cell lines and drugs, edges represent existing cell line-drug associations. Consequently, we computed the heterogeneous graph convolutional features $Cell_f^{(k)}$ of cell lines, as defined by Equation (19) and (20):

$$Z_{cell}^{(k)} = \left(D_c^{-\frac{1}{2}} A D_d^{-\frac{1}{2}} \right) Drug_f^{(k-1)} W_{c1}^{(k-1)} + (D_c^{-1} + I_c) Cell_f^{(k-1)} W_{c2}^{(k-1)} \quad (19)$$

$$Cell_f^{(k)} = \alpha Z_{cell}^{(k)} + \beta (Z_{cell}^{(k)} \odot Cell_f^0) \quad (20)$$

Where, $Cell_f^0$ represent the initial features of cell lines, obtained through Equations (13). The degree matrices of the known cell line-drug association matrix A , thereby $D_{c(ij)} = \sum_j A_{ij} + 1$, $D_{d(ij)} = \sum_i A_{ji} + 1$, where $D_{c(ij)}$ and $D_{d(ij)}$ represents the value at the i -th row and j -th column of matrix

D_c and D_d . W_{c1} and W_{c2} denotes the learnable parameters used in the heterogeneous graph convolution for cell lines. Similarly, the drug features were computed through heterogeneous graph convolution, denoted as $Drug_f^{(k)}$, as described in Equation (21) and (22).

$$Z_{drug}^{(k)} = \left(D_d^{-\frac{1}{2}} A D_c^{-\frac{1}{2}} \right) Cell_f^{(k-1)} W_{d1}^{(k-1)} + (D_d^{-1} + I_d) Drug_f^{(k-1)} W_{d2}^{(k-1)} \quad (21)$$

$$Drug_f^{(k)} = \alpha Z_{drug}^{(k)} + \beta (Z_{drug}^{(k)} \odot Drug_f^0) \quad (22)$$

Here, $Drug_f^0$ represent the initial features of drugs, obtained through Equations (14). W_{d1} and W_{d2} represent the learnable parameters for the heterogeneous graph convolution of drugs.

G. Anticancer Drug Response Prediction

After extracting features of cell lines and drugs from hypergraphs and heterogeneous graphs, we obtained the high-pass features $X_{ch}^{(k)}, X_{dh}^{(k)}$ and low-pass features $X_{cl}^{(k)}, X_{dl}^{(k)}$ from the hypergraph, as well as the convolutional features $Cell_f^{(k)}, Drug_f^{(k)}$ from the heterogeneous graph. We fusion these features to generate the final features for cell lines $Cell_{final}$ and drugs $Drug_{final}$, defined by Equations (23) and (24):

$$Cell_{final} = ReLu \left(X_{ch}^{(k)} + X_{cl}^{(k)} + Cell_f^{(k)} \right) \quad (23)$$

$$Drug_{final} = ReLu \left(X_{dh}^{(k)} + X_{dl}^{(k)} + Drug_f^{(k)} \right) \quad (24)$$

Where, $ReLu$ denotes the activation function in order to introduce non-linearity and enhance feature extraction. The linear correlation coefficient $Corr(h_i, h_j)$ between the final feature of cell lines and drugs was subsequently calculated using Equation (25).

$$Corr(h_i, h_j) = \frac{(h_i - \mu_i)(h_j - \mu_j)^T}{\sqrt{(h_i - \mu_i)(h_i - \mu_i)^T} \sqrt{(h_j - \mu_j)(h_j - \mu_j)^T}} \quad (25)$$

Here, $h_i \in Cell_{final}$ and $h_j \in Drug_{final}$ represent the features of the i th cell line and the j th drug, respectively. Ultimately, the cell line-drug association matrix is reconstructed via equation (26):

$$\hat{A} = \varphi(Corr(Cell_{final}, Drug_{final})) \quad (26)$$

Where, $\varphi(h) = \frac{1}{1+e^{-\gamma h}}$ represents the activation function, and the elements of \hat{A} correspond to the predicted anticancer drug response labels. To optimize the model, we define the following loss function:

$$Loss(A, \hat{A}) = -\frac{1}{m \times n} \sum_{i,j} M_{ij} \left[A_{ij} \ln(\hat{A}_{ij}) + (1 - A_{ij}) \ln(1 - \hat{A}_{ij}) \right] \quad (27)$$

Here, the matrix M represents whether the known cell line-drug association is included in the training set. Specifically, if

the association between cell line i and drug j is part of the training set, then $M_{ij} = 1$. Conversely, $M_{ij} = 0$.

We performed a grid search over the training set with five-fold cross-validation to optimize the model's hyperparameters. The embedding dimensions for cell line gene expressions and drug fingerprints were both determined to be 2040. The Sigmoid scaling factor was set to 8.7, the hyperparameters α and β were set at 0.15 and 0.85. Additionally, the parameter k was fixed at 1. The learning rate was set to 0.001, with a weight decay of 1e-5. The model was trained for 1000 epochs to ensure effective learning and convergence. Algorithm 1 gives the pseudo-code for HRLCDR.

Algorithm 1 HRLCDR method

Input: Drug fingerprint matrix D ; gene expression matrix C ; known cell line-drug response matrix A ; hyperparameters about controlling the relative contributions of feature aggregation and residual connections α and β ; learning rate l_r ; number of epochs e .
Output: Reconstructed the cell line-drug association matrix.

- 1: Calculate the drug similarity matrix S_d by Eq 1.
- 2: Calculate the cell line similarity matrix S_c by Eq 2.
- 3: Building heterogeneous network, drug hypergraph and cell line hypergraph H_d , H_c by S_d , S_c and A .
- 4: **while** trained epochs $< e$ **do**
- 5: Calculate the initial cell line features X_g by Eq 13
- 6: Calculate the initial cell line features X_f by Eq 14
- 7: Calculate the feature of drug and cell lines based on high-pass convolutional layers by Eq 15 and 16
- 8: Calculate the feature of drug and cell lines based on low-pass convolutional layers by Eq 17 and 18
- 9: Calculate the heterogeneous graph convolutional features $Cell_f^{(k)}$ of cell lines by Eq 19 and 20
- 10: Calculate the heterogeneous graph convolutional features $Drug_f^{(k)}$ of cell lines by Eq 21 and 22
- 11: Aggregate cell line feature $Cell_{final}$ and drug feature $Cell_{final}$ about different convolution by Eq 23 and 24
- 12: Reconstruct the cell line-drug association matrix \hat{A} by Eq 26
- 13: Calculate the Loss by Eq 27.
- 14: Optimize model via gradient descent and backpropagation.
- 15: **end while**
- 16: Return \hat{A}

III. EXPERIMENTS

A. Baseline

To evaluate the performance of HRLCDR, we compared it against eight state-of-the-art baselines:

DeepCDR [6] and GraphCDR [8]: Utilizes cell line gene expression and drug molecular graphs as input using different neural network models to predict drug responses (CNN and GCN with contrastive learning, respectively).

DeepDSC [4] and NIHGCN [9]: Uses cell line gene expression and drug molecular fingerprints to predict drug responses. DeepDSC uses an autoencoder for feature extraction, while NIHGCN leverages parallel heterogeneous graph convolution.

MOFGCN [7]: Relies on cell line and drug similarities for drug response prediction using a GCN model.

TSGCNN [10]: Utilizes cell line gene expression and drug molecular fingerprints for response prediction via a dual-space graph convolution.

SubCDR [17]: Use gene expression profiles and drug molecular substructures as input features, leveraging graph

convolutional networks and Singular Value Decomposition, aggregating the information from subcomponent to predict drug responses.

DRPreter [18]: Combines cell line pathway data with drug molecular graphs, while employing a transformer architecture to efficiently diffuse information between cell lines and drugs to predict drug responses.

DeepTTA [19]: The model utilizes gene expression data from cell lines alongside drug substructure data as inputs, and leverages a transformer-based approach to predict drug responses to cell lines.

For fair and consistent comparisons, all baseline methods utilized identical input data—comprising gene expression profiles, drug fingerprints, or molecular graphs. Meanwhile, we adopted the recommended parameter settings from their original studies. Additionally, an independent test set was created by randomly selecting 10% of positive samples along with an equal number of negative samples. The remaining data were used for the cross-validation process, where five five-fold cross-validations were performed on the cross-validation set. This process was repeated five times, and the final results were reported as the average performance across these repetitions. Details regarding these settings can be found in Table I.

B. Experimental Setup

We conducted eight experiments to assess the performance of our model:

Test 1: Hypergraph Combinations: Different hypergraph construction methods were combined to explore the impact of capturing higher-order associations between cell lines and drugs.

Test 2: Parameter Analysis: We performed a parameter sensitivity analysis to explore the influence of different hyperparameters on model performance.

Test 3: Random Zeroing Experiment: We randomly remove a portion of the data to assess the ability of various methods to recover missing known cell line-drug associations.

Test 4: Novel Drug and Cell Line Prediction: By randomly masking 20% of the drug or cell line response data, we evaluate the ability of the models to predict responses to novel drugs or cell lines.

Test 5: IC50 Regression Experiment: We performed regression tests to examine whether HRLCDR's predictions have a stronger correlation with the experimentally measured IC50 compared to the baselines under random zero-cross validation in the GDSC and CCLE datasets.

Test 6: Ablation Study: We analyzed the performance contributions of various convolution operations by removing the low-pass, high-pass, and heterogeneous convolution features.

Test 7: UMAP Visualization: We visualized the learned feature embeddings alongside the original attributes via UMAP to assess the model's discriminative ability.

Test 8: Case Analysis: A set of case studies was conducted to evaluate the model's ability to predict missing drug responses in cell lines.

TABLE I
THE PARAMETERS FOR THE BASELINE METHODS

Algorithm	parameter
DeepCDR	GCN_dim=256,GCN_out=100,cell_out=100,kernel_size=(1,150),kernel_size=(1,5),lr=0.001, weight_decay=0,epoch=1000
DeepDSC	pretrain_dim=[2000,1000,500],drug_dim=256,linear_dim=[1000,800, 500,100,1],lr=0.0004, weight_decay=0,epoch=250,patience=30
MOFGCN	sigma=2,hidden_dim=[192,36],sigmoid_scale=5.74,lr=0.0005,epoch=1000
GraphCDR	hidden_channels=256,output_channels=100,alpha = 0.3,beta = 0.3, lr = 0.001,epoch = 1000
NIHGCN	GCN_dim=1024,alpha=0.25,sigmoid_scale=8,lr=0.001,weight_decay=0.00001,epoch=1000
TSGCNN	scale_parameter=3,kc=7,kd=7,hidden_dim=224,decoder_dim=192,sigmoid_scale=6.9,lr=0.0005,epoch=1000
SubCDR	scale_parameter=1,lr=0.0001,weight_decay=0.0001,epoch=100
DRPreter	drug_dim=128,drug_cell_dim=256,patience=10,hidden_dim=8,lr=0.0001,epoch=300
DeepTTA	MLP_dim=256,transformer_dim=128,classifier_hidden_dims=[1024, 1024, 512],weight_decay=0,lr=0.0001,epoch=300

TABLE II
RESULTS OF RANDOM ZEROING EXPERIMENTS FOR VARIOUS HYPERGRAPH CONSTRUCTION COMBINATIONS

Cell line hypergraph	Drug hypergraph	GDSC		CCLE	
		AUC	AUPRC	AUC	AUPRC
$H_{c1} + H_{c2}$	$H_{d1} + H_{d2}$	$0.87819 \pm 2 \times 10^{-6}$	$0.88274 \pm 2 \times 10^{-6}$	$0.86750 \pm 1 \times 10^{-4}$	$0.86538 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2}$	$H_{d1} + H_{d2} + H_{d3}$	$0.87814 \pm 2 \times 10^{-6}$	$0.88265 \pm 2 \times 10^{-6}$	$0.86844 \pm 1 \times 10^{-4}$	$0.86604 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2}$	$H_{d1} + H_{d2} + H_{d3} + H_{d4}$	$0.87807 \pm 2 \times 10^{-6}$	$0.88264 \pm 2 \times 10^{-6}$	$0.86837 \pm 8 \times 10^{-5}$	$0.86596 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3}$	$H_{d1} + H_{d2}$	$0.87810 \pm 1 \times 10^{-6}$	$0.88258 \pm 2 \times 10^{-6}$	$0.86746 \pm 1 \times 10^{-4}$	$0.86555 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3}$	$H_{d1} + H_{d2} + H_{d3}$	$0.87804 \pm 1 \times 10^{-6}$	$0.88259 \pm 2 \times 10^{-6}$	$0.86811 \pm 9 \times 10^{-5}$	$0.86719 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3}$	$H_{d1} + H_{d2} + H_{d3} + H_{d4}$	$0.87814 \pm 1 \times 10^{-6}$	$0.88262 \pm 2 \times 10^{-6}$	$0.86782 \pm 7 \times 10^{-5}$	$0.86606 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3} + H_{c4}$	$H_{d1} + H_{d2}$	$0.87801 \pm 1 \times 10^{-6}$	$0.88257 \pm 2 \times 10^{-6}$	$0.86884 \pm 8 \times 10^{-5}$	$0.86652 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3} + H_{c4}$	$H_{d1} + H_{d2} + H_{d3}$	$0.87810 \pm 1 \times 10^{-6}$	$0.88266 \pm 2 \times 10^{-6}$	$0.86939 \pm 9 \times 10^{-5}$	$0.86767 \pm 1 \times 10^{-4}$
$H_{c1} + H_{c2} + H_{c3} + H_{c4}$	$H_{d1} + H_{d2} + H_{d3} + H_{d4}$	$0.87804 \pm 2 \times 10^{-6}$	$0.88252 \pm 2 \times 10^{-6}$	$0.86931 \pm 9 \times 10^{-5}$	$0.86725 \pm 1 \times 10^{-4}$

C. Cell-Line and Drug Hypergraphs Construction

To assess the effect of capturing higher-order information from different cell line-drug associations on the experimental outcomes, we conducted random zeroing experiments across various hypergraph construction combinations. In this experiment, $H_{c1}, H_{c2}, H_{c3}, H_{c4}$ and H_{d1}, H_{d2}, H_{d3} were computed using the corresponding formulas, while direct drug-drug associations H_{d4} represented by $H_{d4} = (S_d \cdot A^T) \cdot (S_d \cdot A^T)^T$.

Table II provides a comparative analysis of the performance of various hypergraph-based construction combinations on the GDSC and CCLE datasets. For the GDSC dataset, combination $[H_{c1}, H_{c2}; H_{d1}, H_{d2}]$ achieved the highest performance, with an AUC of 0.87819 and an AUPRC of 0.88274. In contrast, combination $[H_{c1}, H_{c2}, H_{c3}, H_{c4}; H_{d1}, H_{d2}, H_{d3}]$ performed best on the CCLE dataset, yielding an AUPRC of 0.86767 and an AUC of 0.86939. These results underscore the varying effectiveness of hypergraph configurations in modeling complex relationships within the datasets, emphasizing the critical role of selecting the appropriate construct to optimize predictive performance in cancer drug response prediction. As stated previously, our model selects the following structure to construct the hypergraph of cell lines and drugs: The cell line data employed structure $[H_{c1}, H_{c2}, H_{c3}, H_{c4}]$, while the drug data was employed by structure $[H_{d1}, H_{d2}, H_{d3}]$.

D. Parameter Analysis

To investigate the impact of parameters on model performance, we conducted a parameter analysis on the top-K similarity sets of cell lines and drugs during hypergraph construction and on the hyperparameters α and β , which control aggregation and residual weights. Both analyses were performed under random value zeroing experiments. Table III indicates that when the number of similarity sets is small,

certain relevant information may be grouped. However, as the number increases, additional noise may be introduced, negatively affecting the results. Based on this analysis, we set $K = 10$. Table III also reveals that when $\alpha = 0.15$ and $\beta = 0.85$, the CCLE dataset has an AUC value of 0.86939 and an AUPRC value of 0.86767, both of which are the highest. Although the GDSC dataset does not have the highest AUC and AUPRC values, they are still relatively high(AUC of 0.87810 and AUPRC of 0.88266) and are stable compared to other parameter combinations. Ultimately, we set $\alpha = 0.15$ and $\beta = 0.85$.

E. Experiment on Randomly Zeroing Values Cross-Validation

In Test 3, we randomly selected 10% of the positive samples along with an equal number of negative samples to create an independent test set. Additionally, the remaining data were used for the cross-validation process, where five five-fold cross-validations were performed on the cross-validation set. In each round of cross-validation, one-fifth of the positive samples and an equal number of negative samples were designated as the validation set, with the remaining data used for training. The model with the highest AUC on the validation set was selected for prediction. This procedure was repeated five times, and the final results were averaged from the independent test set. Table IV presents the results of Test 3 across all algorithms on the GDSC and CCLE datasets. Our model, HRLCDR, outperforms all baseline methods in terms of classification performance. On the GDSC dataset, HRLCDR achieved an AUC of 87.81% and an AUPRC of 88.27%, surpassing the second-best method by 0.27% and 0.26%, respectively. On the CCLE dataset, HRLCDR achieved an AUC of 86.94% and an AUPRC of 86.77%. Compared to the second-ranked NIHGCN, this represents a 0.90% improvement in AUC and a 0.54%

TABLE III
RESULTS OF THE RANDOM ZEROING EXPERIMENT WITH DIFFERENT PARAMETERS K , α AND β

	AUC	GDSC	AUC	CCLE
	AUPRC	AUPRC	AUPRC	AUPRC
K = 10	$0.87810 \pm 1 \times 10^{-6}$	$0.88266 \pm 2 \times 10^{-6}$	$0.86939 \pm 9 \times 10^{-5}$	$0.86767 \pm 1 \times 10^{-4}$
$K = 5$	$0.87795 \pm 1 \times 10^{-6}$	$0.88238 \pm 2 \times 10^{-6}$	$0.86923 \pm 8 \times 10^{-5}$	$0.86592 \pm 1 \times 10^{-4}$
$K = 15$	$0.87804 \pm 1 \times 10^{-6}$	$0.88257 \pm 2 \times 10^{-6}$	$0.86858 \pm 9 \times 10^{-5}$	$0.86620 \pm 1 \times 10^{-4}$
$K = 20$	$0.87786 \pm 1 \times 10^{-6}$	$0.88234 \pm 2 \times 10^{-6}$	$0.86715 \pm 1 \times 10^{-4}$	$0.86401 \pm 2 \times 10^{-4}$
$\alpha = 0.15, \beta = 0.85$	$0.87810 \pm 1 \times 10^{-6}$	$0.88266 \pm 2 \times 10^{-6}$	$0.86939 \pm 9 \times 10^{-5}$	$0.86767 \pm 1 \times 10^{-4}$
$\alpha = 0.25, \beta = 0.75$	$0.87840 \pm 1 \times 10^{-6}$	$0.88291 \pm 2 \times 10^{-6}$	$0.86859 \pm 8 \times 10^{-5}$	$0.86474 \pm 1 \times 10^{-4}$
$\alpha = 0.2, \beta = 0.8$	$0.87810 \pm 2 \times 10^{-6}$	$0.88255 \pm 2 \times 10^{-6}$	$0.86898 \pm 9 \times 10^{-5}$	$0.86660 \pm 1 \times 10^{-4}$
$\alpha = 0.1, \beta = 0.9$	$0.87786 \pm 2 \times 10^{-6}$	$0.88237 \pm 2 \times 10^{-6}$	$0.86836 \pm 9 \times 10^{-5}$	$0.86494 \pm 1 \times 10^{-4}$
$\alpha = 0.9, \beta = 0.1$	$0.87526 \pm 3 \times 10^{-6}$	$0.87873 \pm 4 \times 10^{-6}$	$0.86466 \pm 8 \times 10^{-5}$	$0.86273 \pm 9 \times 10^{-5}$
$\alpha = 0.85, \beta = 0.15$	$0.87599 \pm 2 \times 10^{-6}$	$0.87950 \pm 4 \times 10^{-6}$	$0.86495 \pm 8 \times 10^{-5}$	$0.86323 \pm 1 \times 10^{-4}$
$\alpha = 0.8, \beta = 0.2$	$0.87697 \pm 2 \times 10^{-6}$	$0.88046 \pm 3 \times 10^{-6}$	$0.86552 \pm 7 \times 10^{-5}$	$0.86374 \pm 1 \times 10^{-4}$
$\alpha = 0.75, \beta = 0.25$	$0.87756 \pm 2 \times 10^{-6}$	$0.88147 \pm 3 \times 10^{-6}$	$0.86462 \pm 8 \times 10^{-5}$	$0.86265 \pm 1 \times 10^{-4}$
$\alpha = 0.5, \beta = 0.5$	$0.87942 \pm 2 \times 10^{-6}$	$0.88339 \pm 3 \times 10^{-6}$	$0.86717 \pm 8 \times 10^{-5}$	$0.86276 \pm 9 \times 10^{-5}$

TABLE IV
COMPARISON OF EVERY METHOD UNDER RANDOMLY ZEROING CROSS-VALIDATION ON TWO DATASETS

Algorithm	GDSC			CCLE
	AUC	AUPRC	AUC	AUPRC
DeepCDR	$0.8030 \pm 3 \times 10^{-3}$	$0.8064 \pm 3 \times 10^{-3}$	$0.8384 \pm 6 \times 10^{-5}$	$0.8292 \pm 2 \times 10^{-4}$
DeepDSC	$0.8083 \pm 7 \times 10^{-5}$	$0.8130 \pm 8 \times 10^{-5}$	$0.8543 \pm 5 \times 10^{-5}$	$0.8489 \pm 5 \times 10^{-5}$
MOFGCN	$0.8685 \pm 3 \times 10^{-6}$	$0.8726 \pm 3 \times 10^{-6}$	$0.8497 \pm 5 \times 10^{-5}$	$0.8514 \pm 7 \times 10^{-5}$
GraphCDR	$0.8176 \pm 6 \times 10^{-5}$	$0.8253 \pm 5 \times 10^{-5}$	$0.8604 \pm 7 \times 10^{-5}$	$0.8623 \pm 6 \times 10^{-5}$
NIHGCN	$0.8754 \pm 3 \times 10^{-6}$	$0.8801 \pm 3 \times 10^{-6}$	$0.8604 \pm 7 \times 10^{-5}$	$0.8623 \pm 6 \times 10^{-5}$
TSGCNN	$0.8706 \pm 3 \times 10^{-6}$	$0.8745 \pm 4 \times 10^{-6}$	$0.8524 \pm 7 \times 10^{-5}$	$0.8542 \pm 6 \times 10^{-5}$
SubCDR	$0.8251 \pm 3 \times 10^{-5}$	$0.8292 \pm 2 \times 10^{-5}$	$0.7802 \pm 7 \times 10^{-5}$	$0.7830 \pm 2 \times 10^{-4}$
DRPreter	$0.8409 \pm 3 \times 10^{-5}$	$0.8473 \pm 3 \times 10^{-5}$	$0.8237 \pm 6 \times 10^{-4}$	$0.8087 \pm 2 \times 10^{-3}$
DeepTTA	$0.8455 \pm 7 \times 10^{-6}$	$0.8494 \pm 7 \times 10^{-6}$	$0.8301 \pm 3 \times 10^{-5}$	$0.8299 \pm 6 \times 10^{-5}$
HRLCDR	$0.8787 \pm 1 \times 10^{-6}$	$0.8827 \pm 2 \times 10^{-6}$	$0.8694 \pm 9 \times 10^{-5}$	$0.8677 \pm 1 \times 10^{-4}$

improvement in AUPRC. In contrast to models like DeepCDR and DeepDSC, which don't focus on the known cell line-drug associations, methods such as MOFGCN, GraphCDR, NIHGCN, TSGCNN, and especially HRLCDR all benefit from exploiting these known associations, enabling more effective learning of predictive representations. HRLCDR outperforms other graph convolution-based models (e.g., NIHGCN and GraphCDR) by integrating not only first-order relationships but also hypergraph learning. The hypergraph structure enables HRLCDR to capture higher-order relationships between nodes, improving feature aggregation and resulting in enhanced cell line and drug feature representations.

F. Prediction Experiments for Novel Drugs and Cell Lines

To evaluate how well HRLCDR predicts responses to new drugs and cell lines, we conducted Test 4 on the GDSC and CCLE datasets. For novel drugs/cell lines, we randomly removed one-fifth of the columns/rows from the cell line-drug association matrix. All positive samples from the removed columns, along with an equal number of negative samples, were used for testing. The remaining association matrix served as the five-fold cross-validation set. In each fold, we chose one-fifth of the remaining positive samples and an equal number of randomly selected negative samples from the cross-validation set for the validation. The remaining positive and negative were used for training. The model with the best AUC results on the validation set was used to predict response of the test set. The process was repeated until all columns had

been tested. Finally, the average results across all test sets were reported.

To verify the algorithm accurately predicts the responses of new drugs, we removed one-fifth of the columns from the cell line-drug association matrix. Table V summarizes the performance of all algorithms in predicting reactions to novel drugs on the GDSC and CCLE datasets. HRLCDR achieved the highest performance on the CCLE dataset, attaining the best AUC and second-best AUPRC on the GDSC dataset. In the GDSC dataset, our model achieved an AUC of 70.25% and an AUPRC of 68.76% (with improvements of 0.17% AUC over the second-ranked method). In the CCLE dataset, our model had an AUC of 71.62% and an AUPRC of 69.59% (improvements of 5.45% AUC and 5.31% AUPRC over the second-ranked method). The results indicate that where no known associations between cell lines and drugs were available, hypergraphs incorporate higher-order relationships between nodes, allowing HRLCDR to capture more complex interactions between cell lines and drugs.

To verify the algorithm accurately predicts the responses of new cell lines, we removed one-fifth of the rows from the cell line-drug association matrix. Table V reports the performance of all algorithms in predicting responses for new cell lines on the GDSC and CCLE datasets. Our model maintained its strong performance in this test. Our model achieved an AUC of 77.01% and an AUPRC of 78.00% in the GDSC dataset, surpassing the second-best method of 0.12% and 0.11%, respectively. Similarly, in the CCLE dataset, HRLCDR achieved an AUC of 78.68% and an AUPRC of 77.04%, outperforming

TABLE V
RESULTS OF NOVEL DRUGS AND CELL LINES RESPONSE PREDICTION ON TWO DATASETS

Algorithm	GDSC new cell lines		CCLE new cell lines		GDSC new drugs		CCLE new drugs	
	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
DeepCDR	$0.6221 \pm 7 \times 10^{-3}$	$0.6210 \pm 7 \times 10^{-3}$	$0.6542 \pm 1 \times 10^{-2}$	$0.6402 \pm 8 \times 10^{-3}$	$0.6918 \pm 4 \times 10^{-4}$	$0.7008 \pm 5 \times 10^{-4}$	$0.6736 \pm 8 \times 10^{-3}$	$0.6467 \pm 6 \times 10^{-3}$
DeepDSC	$0.6421 \pm 7 \times 10^{-3}$	$0.6476 \pm 8 \times 10^{-3}$	$0.5745 \pm 1 \times 10^{-2}$	$0.5682 \pm 9 \times 10^{-3}$	$0.6592 \pm 8 \times 10^{-3}$	$0.6640 \pm 9 \times 10^{-3}$	$0.6108 \pm 2 \times 10^{-2}$	$0.6074 \pm 2 \times 10^{-2}$
MOFGCN	$0.6845 \pm 3 \times 10^{-4}$	$0.6817 \pm 3 \times 10^{-4}$	$0.6526 \pm 1 \times 10^{-3}$	$0.6423 \pm 1 \times 10^{-3}$	$0.5852 \pm 1 \times 10^{-4}$	$0.5969 \pm 1 \times 10^{-4}$	$0.6712 \pm 5 \times 10^{-4}$	$0.6230 \pm 5 \times 10^{-4}$
GraphCDR	$0.6494 \pm 3 \times 10^{-4}$	$0.6003 \pm 2 \times 10^{-4}$	$0.5731 \pm 3 \times 10^{-3}$	$0.5465 \pm 1 \times 10^{-3}$	$0.6946 \pm 3 \times 10^{-4}$	$0.6792 \pm 3 \times 10^{-4}$	$0.6330 \pm 7 \times 10^{-4}$	$0.6088 \pm 1 \times 10^{-3}$
NIHGCN	$0.6861 \pm 8 \times 10^{-5}$	$0.6709 \pm 3 \times 10^{-4}$	$0.6381 \pm 1 \times 10^{-2}$	$0.6298 \pm 9 \times 10^{-3}$	$0.7689 \pm 3 \times 10^{-5}$	$0.7721 \pm 9 \times 10^{-5}$	$0.7663 \pm 3 \times 10^{-4}$	$0.7551 \pm 3 \times 10^{-4}$
TSGCNN	$0.6593 \pm 5 \times 10^{-4}$	$0.6600 \pm 5 \times 10^{-4}$	$0.6617 \pm 3 \times 10^{-3}$	$0.6428 \pm 2 \times 10^{-3}$	$0.5964 \pm 1 \times 10^{-4}$	$0.6179 \pm 1 \times 10^{-4}$	$0.6440 \pm 1 \times 10^{-3}$	$0.6199 \pm 6 \times 10^{-4}$
SubCDR	$0.6536 \pm 7 \times 10^{-4}$	$0.6536 \pm 3 \times 10^{-4}$	$0.6162 \pm 2 \times 10^{-3}$	$0.6120 \pm 1 \times 10^{-3}$	$0.5964 \pm 1 \times 10^{-4}$	$0.6179 \pm 1 \times 10^{-4}$	$0.6440 \pm 1 \times 10^{-3}$	$0.6199 \pm 6 \times 10^{-4}$
DRPreter	$0.7008 \pm 4 \times 10^{-4}$	$0.6963 \pm 4 \times 10^{-4}$	$0.6484 \pm 5 \times 10^{-3}$	$0.6291 \pm 4 \times 10^{-3}$	$0.7487 \pm 2 \times 10^{-4}$	$0.7631 \pm 1 \times 10^{-4}$	$0.7586 \pm 8 \times 10^{-4}$	$0.7327 \pm 1 \times 10^{-3}$
DeepTTA	$0.6956 \pm 2 \times 10^{-4}$	$0.6872 \pm 2 \times 10^{-4}$	$0.6566 \pm 4 \times 10^{-3}$	$0.6349 \pm 3 \times 10^{-3}$	$0.7692 \pm 8 \times 10^{-5}$	$0.7789 \pm 6 \times 10^{-5}$	$0.7859 \pm 3 \times 10^{-4}$	$0.7648 \pm 4 \times 10^{-4}$
HRLCDR	$0.7025 \pm 3 \times 10^{-4}$	$0.6876 \pm 2 \times 10^{-4}$	$0.7162 \pm 3 \times 10^{-3}$	$0.6959 \pm 3 \times 10^{-3}$	$0.7701 \pm 2 \times 10^{-4}$	$0.7800 \pm 2 \times 10^{-4}$	$0.7868 \pm 5 \times 10^{-4}$	$0.7704 \pm 8 \times 10^{-4}$

the second-best method of 0.09% and 0.56%, respectively. These results highlight the effectiveness of using hypergraphs to capture higher-order interactions between drugs and cell lines. By incorporating this information, HRLCDR can learn more complex relationships, leading to superior prediction accuracy for entirely new cell lines.

G. Regression test

We implemented regression test to assess how well the model predictions correlate with the experimentally measured half-maximal inhibitory concentration (IC50) values. The evaluation process is similar to Test 3. We employed three regression metrics: Pearson Correlation Coefficient (PCC), Spearman's Correlation Coefficient (SCC), and Root Mean Square Error (RMSE) to assess the predictive performance of various models. We adjusted the model's output of Equation 25 to predict the IC50 values. Table VI illustrates the correlation between model predictions and experimentally measured IC50 values. HRLCDR achieved the best performance on both the CCLE and GDSC datasets according to all three metrics. Specifically, in the CCLE dataset, its PCC reached 0.6722, and SCC attained 0.7191, surpassing the second-ranked NIHGCN by 0.0055 and 0.142, respectively. The RMSE was 2.6539, a reduction of 0.1054 compared to the second-best method. In the GDSC dataset, the HRLCDR had a PCC of 0.9280, an SCC of 0.9151, and an RMSE of 1.2089. The experimental results from both datasets indicate that our model has a high SCC score, suggesting it effectively captures non-linear trends in drug response. Moreover, our method leading performance in both PCC and RMSE metrics on the CCLE dataset indicates that HRLCDR's predictions are not only accurate in terms of ranking potential drugs but also achieve good precision in predicting actual IC50 values.

H. Ablation Study

Our model is designed to effectively learn the features of cell lines and drugs to predict cell line responses to drugs. To evaluate the performance contributions of different convolutional components, we developed the following model variants by selectively removing specific convolutional features. We conducted comparisons under a random value zeroing experiment. HRLCDR-L: Removing low-pass convolutions; HRLCDR-H: Removing high-pass convolutions;

HRLCDR-D: Removing heterogeneous graph convolutions; HRLCDR-HD: Removing both high-pass convolutions and heterogeneous graph convolutions; HRLCDR-LD: Removing both low-pass convolutions and heterogeneous graph convolutions; HRLCDR-LH: Removing both low-pass convolutions and high-pass convolutions.

Based on the experimental results in Table VII, the model HRLCDR achieves the best performance on the GDSC dataset, with an AUC of 0.87810 and an AUPRC of 0.88266, indicating that all convolutional features contribute to the model's overall performance. When high-pass convolutions (HRLCDR-H) and heterogeneous graph convolutions (HRLCDR-D) are removed, both AUC and AUPRC decrease slightly, suggesting that these components play a crucial role in enhancing performance. On the CCLE dataset, HRLCDR-H achieves the highest AUC (0.86961), while HRLCDR-HD achieves the highest AUPRC (0.86888). These results indicate that high-pass convolutions significantly contribute to AUC, whereas heterogeneous graph convolutions are essential for improving AUPRC. Overall, the removal of individual or combined convolutional components leads to performance degradation, highlighting the complementary roles of different convolution types in the model.

I. UMAP Visualization

We evaluated the performance of the HRLCDR model in predicting responses to CDK4-targeted drugs using compounds obtained from the GDSC and CCLE databases. CDK4, a cyclin-dependent kinase, promotes cell proliferation by interacting with proteins that inhibit retinoblastoma [20]. From the GDSC database, we selected three targeted drugs: PD-0332991, AT-7519, and CGP-082996. From the CCLE database, we included one targeted drug: PD-0332991. To assess model performance, we excluded the corresponding column for one targeted drug in the cell line-drug association matrix, designating it as the test set, and trained the model on the remaining data. This approach facilitated the investigation of drug responses and the identification of optimal cell line features for the test set.

On the one hand, the embedding features learned by the HRLCDR model, along with the original features, were clustered into two groups using the k-means clustering method. Internal validation was conducted using metrics such as the Silhouette Coefficient (SC), Davies-Bouldin Index (DBI), and

TABLE VI
REGRESSION EXPERIMENT RESULTS

Algorithm	GDSC		RMSE	PCC	CCLE		RMSE
	PCC	SCC			SCC	CCL	
DeepCDR	$0.8507 \pm 1 \times 10^{-3}$	$0.8238 \pm 2 \times 10^{-3}$	$1.7175 \pm 5 \times 10^{-2}$	$0.6157 \pm 3 \times 10^{-4}$	$0.6546 \pm 3 \times 10^{-4}$	$3.0789 \pm 5 \times 10^{-3}$	
DeepDSC	$0.8616 \pm 3 \times 10^{-5}$	$0.8341 \pm 4 \times 10^{-5}$	$1.5708 \pm 4 \times 10^{-3}$	$0.5993 \pm 7 \times 10^{-5}$	$0.6294 \pm 8 \times 10^{-5}$	$2.9409 \pm 2 \times 10^{-2}$	
MOFGCN	$0.9103 \pm 7 \times 10^{-8}$	$0.8983 \pm 3 \times 10^{-7}$	$1.3352 \pm 3 \times 10^{-6}$	$0.4959 \pm 2 \times 10^{-5}$	$0.5846 \pm 5 \times 10^{-5}$	$2.9888 \pm 7 \times 10^{-5}$	
GraphCDR	$0.8447 \pm 4 \times 10^{-5}$	$0.8239 \pm 8 \times 10^{-5}$	$1.7230 \pm 1 \times 10^{-3}$	$0.5783 \pm 3 \times 10^{-4}$	$0.6121 \pm 7 \times 10^{-4}$	$3.2174 \pm 7 \times 10^{-3}$	
NIHGCN	$0.9266 \pm 2 \times 10^{-7}$	$0.9143 \pm 3 \times 10^{-7}$	$1.2276 \pm 3 \times 10^{-5}$	$0.6569 \pm 7 \times 10^{-5}$	$0.7049 \pm 1 \times 10^{-4}$	$2.7593 \pm 1 \times 10^{-3}$	
TSGCNN	$0.9210 \pm 8 \times 10^{-8}$	$0.9081 \pm 2 \times 10^{-7}$	$1.2334 \pm 4 \times 10^{-4}$	$0.5915 \pm 3 \times 10^{-5}$	$0.6681 \pm 1 \times 10^{-4}$	$2.8719 \pm 3 \times 10^{-4}$	
SubCDR	$0.9076 \pm 7 \times 10^{-6}$	$0.8887 \pm 1 \times 10^{-5}$	$1.2660 \pm 3 \times 10^{-3}$	$0.6697 \pm 6 \times 10^{-5}$	$0.6965 \pm 2 \times 10^{-4}$	2.6498 $\pm 4 \times 10^{-3}$	
DRPreter	0.9309 $\pm 1 \times 10^{-5}$	$0.9113 \pm 2 \times 10^{-5}$	1.1382 $\pm 1 \times 10^{-3}$	$0.6667 \pm 1 \times 10^{-3}$	$0.6832 \pm 1 \times 10^{-3}$	$2.7939 \pm 3 \times 10^{-2}$	
DeepTTA	$0.9136 \pm 2 \times 10^{-6}$	$0.8894 \pm 2 \times 10^{-6}$	$1.2147 \pm 1 \times 10^{-4}$	$0.5776 \pm 8 \times 10^{-5}$	$0.5872 \pm 2 \times 10^{-4}$	$2.8774 \pm 4 \times 10^{-3}$	
HRLCDR	$0.9280 \pm 4 \times 10^{-7}$	0.9151 $\pm 7 \times 10^{-7}$	$1.2089 \pm 7 \times 10^{-4}$	0.6722 $\pm 3 \times 10^{-5}$	0.7191 $\pm 7 \times 10^{-5}$	$2.6539 \pm 7 \times 10^{-4}$	

TABLE VII
ABLATION EXPERIMENTS ON GDSC AND CCLE DATASETS

Algorithm	GDSC		CCLE	
	AUC	AUPRC	AUC	AUPRC
HRLCDR	$0.87810 \pm 1 \times 10^{-6}$	$0.88266 \pm 2 \times 10^{-6}$	$0.86939 \pm 9 \times 10^{-5}$	$0.86767 \pm 1 \times 10^{-4}$
HRLCDR-L	$0.87806 \pm 2 \times 10^{-6}$	$0.88253 \pm 2 \times 10^{-6}$	$0.86909 \pm 8 \times 10^{-5}$	$0.86724 \pm 1 \times 10^{-4}$
HRLCDR-H	$0.87802 \pm 2 \times 10^{-6}$	$0.88257 \pm 2 \times 10^{-6}$	$0.86961 \pm 9 \times 10^{-5}$	$0.86695 \pm 1 \times 10^{-4}$
HRLCDR-D	$0.87796 \pm 1 \times 10^{-6}$	$0.88242 \pm 2 \times 10^{-6}$	$0.86904 \pm 8 \times 10^{-5}$	$0.86697 \pm 1 \times 10^{-4}$
HRLCDR-HD	$0.87606 \pm 3 \times 10^{-6}$	$0.88101 \pm 3 \times 10^{-6}$	$0.86926 \pm 9 \times 10^{-5}$	$0.86888 \pm 1 \times 10^{-4}$
HRLCDR-LD	$0.87794 \pm 2 \times 10^{-6}$	$0.88249 \pm 2 \times 10^{-6}$	$0.86867 \pm 8 \times 10^{-5}$	$0.86603 \pm 1 \times 10^{-4}$
HRLCDR-LH	$0.87799 \pm 2 \times 10^{-6}$	$0.88255 \pm 2 \times 10^{-6}$	$0.86888 \pm 8 \times 10^{-5}$	$0.86598 \pm 1 \times 10^{-4}$

Normalized Mutual Information (NMI). Table VIII demonstrates that the embedding features effectively separated the cell lines into two groups, achieving high intra-cluster compactness and inter-cluster separability. Additionally, the high NMI score indicates that the embedding features accurately distinguished sensitive and resistant cell lines, demonstrating substantial agreement with the benchmark dataset.

TABLE VIII
CELL LINE CLUSTERING RESULTS: ORIGINAL & LEARNED FEATURES

Datasets	Target drugs	Sihouette	DBI	NMI
GDSC	PD-0332991	Initial features	0.1211	2.7665
		Learned embeddings	0.6395	0.7378 0.0331
	AT-7519	Initial features	0.1161	2.8970
		Learned embeddings	0.6669	0.6648 0.3473
	CGP-082996	Initial features	0.1287	2.4121
		Learned embeddings	0.5090	1.0303 0.0670
CCLE	PD-0332991	Initial features	0.1656	2.1033
		Learned embeddings	0.5143	1.4901 0.0354

On the other hand, we utilized the UMAP tool for visualization. Figure 2 indicates that for the three drugs from the GDSC database, our model effectively classified the cell lines into two distinct groups, outperforming the original features. However, in the case of CCLE: PD-0332991, the separation was less pronounced, potentially due to the blurring of high-dimensional features during dimensionality reduction by UMAP. Overall, the results suggest that our model exhibits strong discriminative capability.

J. Case Analysis

Previous studies have indicated that approximately 20% of drug responses are missing from the Genomics of Drug Sensitivity in Cancer (GDSC) dataset [3]. To evaluate the performance of our model in predicting novel drug responses

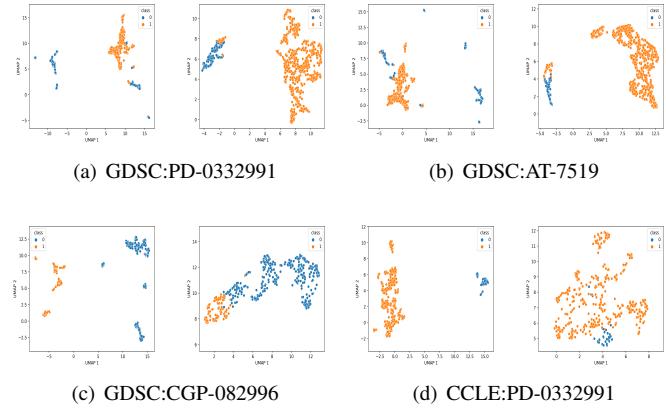


Fig. 2. UMAP visualization of cell line clustering based on original features or learned embeddings, where the left image is initial features, and the right image is learned embeddings.

across cell lines, we trained it on all of the known drug response data from the GDSC dataset. Subsequently, the model's ability to predict the unidentified drug responses within the GDSC dataset was assessed. Specifically, we evaluate two clinically approved drugs: GSK690693 and Bortezomib. We chose these two drugs based on their clinical significance and the robust prior experimental evidence supporting their mechanisms of action. GSK690693 is an Akt kinase inhibitor, and the Akt kinase targets a critical regulator of cell survival, metabolism, proliferation, and apoptosis. At the same time, Bortezomib is a proteasome inhibitor used in the treatment of multiple myeloma and various types of lymphoma. Table IX presents the top ten cell lines predicted by our model to be sensitive to GSK690693 and Bortezomib. Among those sensitive to GSK690693, the model identified cell lines such

as RCH-ACV, MOLT-16, and JEKO-1. Notably, [21] reported that GSK690693 effectively inhibited proliferation in the pre-B cell line RCH-ACV and the T cell line MOLT-16, suggesting its therapeutic potential in these cell types. Likewise, [22] found that GSK690693 suppressed proliferation in the mantle cell lymphoma (MCL) line JEKO-1. Our model predicted that Bortezomib is sensitive to cell lines A375, AGS, HOS, MEL-JUSO, MIAPaCa-2, HGC-27, and C32. [23] confirmed that Bortezomib exhibited potent anti-proliferative effects on the melanoma cell line A375. [24] demonstrated that Bortezomib downregulates substrate degradation in AGS cells, promoting apoptosis. [25] further observed that Bortezomib inhibited growth in HOS cells in a dose- and time-dependent manner, triggering autophagy and apoptosis, processes associated with changes in MAPK signaling. [26] highlighted the role of BCL-2 expression in MIAPaCa-2 cells, noting that Bortezomib-induced inhibition of proteasomal activity reduces BCL-2 levels via NF- κ B suppression, thereby enhancing apoptotic responses. [27] indicated that Bortezomib showed significant cytotoxicity in primary chordoma cells, C24, C25 and C32, with modest growth inhibitor activity in most of these cells. While [28] demonstrate that Bortezomib has significant inhibitory effects on the growth of human gastric cancer cells HGC-27, and the Bortezomib significantly decreased the viability of HGC-27 cells and induced apoptosis. These results collectively validate the predictive capacity of our model and underscore its potential for identifying cell lines sensitive to specific drugs, advancing personalized cancer therapy.

TABLE IX
TOP 10 PREDICTED SENSITIVE CELL LINES FOR NOVEL DRUGS

Drug	Prediction	Rank	Cell line	PMID
GSK690693	Sensitive	1	CRO-AP2	N/A
		2	DOHH-2	N/A
		3	GA-10	N/A
		4	JEKO-1	32120074
		5	KP-1N	N/A
		6	MM1S	N/A
		7	MOLT-16	19064730
		8	NCI-H929	N/A
		9	RCH-ACV	19064730
		10	SK-GT-2	N/A
Bortezomib	Sensitive	1	A375	27677689
		2	AGS	29843100
		3	C32	23792643
		4	COLO-679	N/A
		5	HGC-27	25602436
		6	HOS	23975859
		7	Hs-633T	N/A
		8	KON	N/A
		9	MEL-JUSO	N/A
		10	MIAPaCa-2	15791457

IV. CONCLUSION

This paper introduces HRLCDR, an innovative model for predicting drug response that leverages hypergraph representation learning. Unlike traditional approaches, HRLCDR utilizes both low-pass and high-pass hypergraph convolutions to capture higher-order interactions between cell lines and drugs, thereby enriching the representation of cell line–drug relationships. Additionally, the model incorporates features

from a heterogeneous graph, further enhancing the modeling of complex dependencies between these entities. By combining these two complementary feature sets, HRLCDR effectively reconstructs cell line–drug associations, leading to more accurate drug response predictions. The experimental results demonstrate that HRLCDR consistently outperforms existing methods, achieving superior predictive accuracy in drug response prediction tasks. However, this study has several limitations. Simple direct connections for feature integration may overlook complex relationships, and relying on a single type of hypergraph convolution could limit interaction modeling. Scalability and generalizability also remain challenges, especially with large-scale data. Future work will focus on improving hypergraph computation techniques, integrating multiple omics data types, and addressing the scalability and generalizability challenges to enhance HRLCDR’s broader applicability across various cancer types and drug classes.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China (No. 62472202, No.61972185). Yunnan Ten Thousand Talents Plan young.

REFERENCES

- [1] H. S. Chan, H. Shan, T. Dahoun, H. Vogel, and S. Yuan, “Advancing drug discovery via artificial intelligence,” *Trends in pharmacological sciences*, vol. 40, no. 8, pp. 592–604, 2019.
- [2] M. Ghandi, F. W. Huang, J. Jané-Valbuena, G. V. Kryukov, C. C. Lo, E. R. McDonald III, J. Barretina, E. T. Gelfand, C. M. Bielski, H. Li *et al.*, “Next-generation characterization of the cancer cell line encyclopedia,” *Nature*, vol. 569, no. 7757, pp. 503–508, 2019.
- [3] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson *et al.*, “Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells,” *Nucleic acids research*, vol. 41, no. D1, pp. D955–D961, 2012.
- [4] M. Li, Y. Wang, R. Zheng, X. Shi, Y. Li, F.-X. Wu, and J. Wang, “Deepdsc: a deep learning method to predict drug sensitivity of cancer cell lines,” *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 2, pp. 575–582, 2019.
- [5] R. Su, X. Liu, L. Wei, and Q. Zou, “Deep-resp-forest: a deep forest model to predict anti-cancer drug response,” *Methods*, vol. 166, pp. 91–102, 2019.
- [6] Q. Liu, Z. Hu, R. Jiang, and M. Zhou, “Deepcdr: a hybrid graph convolutional network for predicting cancer drug response,” *Bioinformatics*, vol. 36, no. Supplement_2, pp. i911–i918, 2020.
- [7] W. Peng, T. Chen, and W. Dai, “Predicting drug response based on multi-omics fusion and graph convolution,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1384–1393, 2021.
- [8] X. Liu, C. Song, F. Huang, H. Fu, W. Xiao, and W. Zhang, “Graphcdr: a graph neural network method with contrastive learning for cancer drug response prediction,” *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab457, 2022.
- [9] W. Peng, H. Liu, W. Dai, N. Yu, and J. Wang, “Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions,” *Bioinformatics*, vol. 38, no. 19, pp. 4546–4553, 2022.
- [10] W. Peng, T. Chen, H. Liu, W. Dai, N. Yu, and W. Lan, “Improving drug response prediction based on two-space graph convolution,” *Computers in Biology and Medicine*, vol. 158, p. 106859, 2023.
- [11] M. Chen, Y. Jiang, X. Lei, Y. Pan, C. Ji, W. Jiang, and H. Xiong, “Drug-target interactions prediction based on signed heterogeneous graph neural networks,” *Chinese Journal of Electronics*, vol. 33, no. 1, pp. 231–244, 2024.
- [12] W. Peng, Z. He, W. Dai, and W. Lan, “Mhclmda: multihypergraph contrastive learning for mirna–disease association prediction,” *Briefings in Bioinformatics*, vol. 25, no. 1, p. bbad524, 2024.

- [13] S. Jin, Y. Hong, L. Zeng, Y. Jiang, Y. Lin, L. Wei, Z. Yu, X. Zeng, and X. Liu, "A general hypergraph learning algorithm for drug multi-task predictions in micro-to-macro biomedical networks," *PLOS Computational Biology*, vol. 19, no. 11, p. e1011597, 2023.
- [14] A. Bretto, "Hypergraph theory," *An introduction. Mathematical Engineering*. Cham: Springer, vol. 1, 2013.
- [15] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.
- [16] A. Bojchevski, J. Gasteiger, B. Perozzi, A. Kapoor, M. Blais, B. Różemberczki, M. Lukasik, and S. Günnemann, "Scaling graph neural networks with approximate pagerank," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 2464–2473.
- [17] X. Liu and W. Zhang, "A subcomponent-guided deep learning method for interpretable cancer drug response prediction," *PLOS Computational Biology*, vol. 19, no. 8, p. e1011382, 2023.
- [18] J. Shin, Y. Piao, D. Bang, S. Kim, and K. Jo, "Drpreter: interpretable anticancer drug response prediction using knowledge-guided graph neural networks and transformer," *International Journal of Molecular Sciences*, vol. 23, no. 22, p. 13919, 2022.
- [19] L. Jiang, C. Jiang, X. Yu, R. Fu, S. Jin, and X. Liu, "Deeppta: a transformer-based model for predicting cancer drug response," *Briefings in bioinformatics*, vol. 23, no. 3, p. bbac100, 2022.
- [20] J. E. Staunton, D. K. Slonim, H. A. Coller, P. Tamayo, M. J. Angelo, J. Park, U. Scherf, J. K. Lee, W. O. Reinhold, J. N. Weinstein *et al.*, "Chemosensitivity prediction by transcriptional profiling," *Proceedings of the National Academy of Sciences*, vol. 98, no. 19, pp. 10787–10792, 2001.
- [21] D. S. Levy, J. A. Kahana, and R. Kumar, "Akt inhibitor, gsk690693, induces growth inhibition and apoptosis in acute lymphoblastic leukemia cell lines," *Blood, The Journal of the American Society of Hematology*, vol. 113, no. 8, pp. 1723–1729, 2009.
- [22] Y. Liu, Z. Zhang, F. Ran, K. Guo, X. Chen, and G. Zhao, "Extensive investigation of benzylic n-containing substituents on the pyrrolopyrimidine skeleton as akt inhibitors with potent anticancer activity," *Bioorganic Chemistry*, vol. 97, p. 103671, 2020.
- [23] U. A Rossi, L. ME Finocchiaro, and G. C Gilkin, "Bortezomib enhances the antitumor effects of interferon- β gene transfer on melanoma cells," *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, vol. 17, no. 5, pp. 754–761, 2017.
- [24] P. Lu, Y. Guo, L. Zhu, Y. Xia, Y. Zhong, and Y. Wang, "A novel nac/uae dual inhibitor Ip0040 blocks neddylation and ubiquitination leading to growth inhibition and apoptosis of cancer cells," *European Journal of Medicinal Chemistry*, vol. 154, pp. 294–304, 2018.
- [25] Z. Lou, T. Ren, X. Peng, Y. Sun, G. Jiao, Q. Lu, S. Zhang, X. Lu, and W. Guo, "Bortezomib induces apoptosis and autophagy in osteosarcoma cells through mitogen-activated protein kinase pathway in vitro," *Journal of international medical research*, vol. 41, no. 5, pp. 1505–1519, 2013.
- [26] B. N. Fahy, M. G. Schlieman, M. M. Mortenson, S. Virudachalam, and R. J. Bold, "Targeting bcl-2 overexpression in various human malignancies through nf- κ b inhibition by the proteasome inhibitor bortezomib," *Cancer chemotherapy and pharmacology*, vol. 56, pp. 46–54, 2005.
- [27] M. Xia, R. Huang, S. Sakamuru, D. Alcorta, M.-H. Cho, D.-H. Lee, D. M. Park, M. J. Kelley, J. Sommer, and C. P. Austin, "Identification of repurposed small molecule drugs for chordoma therapy," *Cancer biology & therapy*, vol. 14, no. 7, pp. 638–647, 2013.
- [28] B. Zhang and Y. Gu, "Bortezomib inhibits gastric carcinoma hgc-27 cells through the phospho-jun n-terminal kinase (p-jnk) pathway in vitro," *Gene*, vol. 559, no. 2, pp. 164–171, 2015.



Wei Peng received the PhD degree in computer science from Central South University, China, in 2013. Currently, she is a Professor of Kunming University of Science and Technology, China. Her research interests include bioinformatics and data mining.



Xinyue Xu received the B.E degree from Kunming University of Science and Technology in 2022. She is a master student in Kunming University of Science and Technology, China. Her research interests include bioinformatics and medical image processing.



Jiangzhen Lin is a master student at Kunming University of Science and Technology, China. His research interests include bioinformatics and data mining.



Gong Chen received the B.E degree from Kunming University of Science and Technology in 2023. He is a master student in Kunming University of Science and Technology, China. His research interests include bioinformatics and feature extraction.



Wei Dai received his PhD degree in Computer Application from the University of Chinese Academy of Sciences, China, in 2018. Currently, he is an Associate Professor at Kunming University of Science and Technology. His research interests include bioinformatics, distributed and cloud computing, and data mining.



Xiaodong Fu received his PhD degree from the University of Chinese Academy of Sciences, China, in 2008. Currently, he is a Professor at Kunming University of Science and Technology. His research interests include Service Computing, Intelligent Decision-Making, Big Data, and Software Engineering.



Li Liu received his PhD degree from the Sun Yat-sen University, China, in 2014. Currently, she is a Professor at Kunming University of Science and Technology. Her research interests include Video Image Processing, Multimedia, Computer Vision, Computer Aided Design and Graphics.



Lijun Liu received his PhD degree from the Yunnan University, China, in 2021. Currently, he is an Associate Professor at Kunming University of Science and Technology. His research interests include Medical imaging intelligent diagnosis and report generation, medical vision Q&A, medical vision dialogue and radiology knowledge graph.



Ning Yu Ning Yu works as an Associate Professor in the Department of Computing Sciences, State University of New York Brockport, NY, USA. His current research focuses on artificial intelligence, big data mining and analysis, and high performance computing.