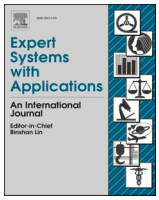
Expert Systems With Applications 297 (2026) 129320



Contents lists available at ScienceDirect

Expert Systems With Applications

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



HCLGT-DRP: Hypergraph contrastive learning and graph transformer for drug response prediction

Wen Zhang [a], Hang Qiu [a,b,*]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China

ARTICLE INFO

Keywords:
Drug response prediction
Hypergraph learning
Graph transformer
Multi-omics
Contrastive learning

ABSTRACT

Accurate drug response prediction (DRP) is crucial for personalized therapy. Recently, graph representation learning has demonstrated promising potential in DRP due to its robust expressive capability. However, existing graph-based methods still face three major issues: 1) failing to capture the high-order complex relations among cell lines, 2) not fully utilizing drug structure information, and 3) insufficiently leveraging cell line-drug association. To address these issues, a novel computational framework, termed HCLGT-DRP, is proposed based on hypergraph contrastive learning and graph transformer for predicting drug response. Specifically, we construct a hypergraph for each type of omics and introduce a hypergraph aggregation mechanism to optimize the hypergraph structure, enabling the capture of high-order relations within omics-specific hypergraphs. To explore interactions across multiple omics data, a hypergraph fusion approach based on a graph-level attention mechanism is proposed to integrate omics-specific hypergraphs into a unified one, enhancing the representation of complex multi-relational data. Moreover, a contrastive learning approach is employed to enhance the ability of representation learning across hypergraphs for multi-omics and heterogeneous graphs for cell line-drug association, obtaining the effective representations of cell lines. Furthermore, a graph transformer with molecular-specific attention is designed to learn drug representations, using the structural information within drug molecules. Extensive experiments conducted on two publicly available datasets show that HCLGT-DRP surpasses state-of-the-art methods across various DRP tasks. External validation on *in vivo* dataset further confirms its effectiveness. This study reveals the value of hypergraph structures in capturing high-order relations among cell lines while emphasizing the potential of sufficiently utilizing drug structure information and cell line-drug association to enhance the predictive performance.

1. Introduction

Cancer persists as a foremost contributor of death worldwide (Cancer, n.d.) representing a serious threat to human health. Due to differences in genome profiles among cancer patients, individuals with the same type of cancer often exhibit varied responses to identical drug treatments (Garnett et al., 2012), presenting significant challenges for therapy. Therefore, accurately predicting individual patient response to specific drug is essential for achieving personalized medicine, potentially improving therapeutic outcomes and reducing expenses (Baptista et al., 2021).

Breakthroughs in high-throughput sequencing technologies, coupled with the establishment of preclinical anti-cancer drug datasets, such as the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2012) and the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), have generated drug response data and comprehensive multi-omics profiles for cell lines, such as genomic, transcriptomic, epigenetic. The large public repository, PubChem(Kim et al., 2019), provides various chemical and structure information of drug molecules, such as molecular graphs, SMILES strings, and fingerprints. These data have substantially facilitated the adoption of deep learning (DL) methods in DRP (Lin et al., 2024). For example, Hossein et al. (Sharifi-Noghabi et al., 2019) introduced a deep neural network approach termed MOLI, which employs distinct encoding sub-networks for single-omics data and leverages a classification sub-network to predict drug response. Min et al. (M. Li et al., 2021) integrated genomic features of cancer cell lines using

* Corresponding author at: School of Computer Science and Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, Sichuan 611731, China.
E-mail addresses: wenz@std.uestc.edu.cn (W. Zhang), qiuhang@uestc.edu.cn (H. Qiu).

https://doi.org/10.1016/j.eswa.2025.129320
Received 21 January 2025; Received in revised form 3 July 2025; Accepted 8 August 2025
Available online 9 August 2025
0957-4174/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

a stacked deep autoencoder and combined them with chemical information of compounds to model drug sensitivity. However, traditional DL methods rely on handcrafted features or predefined descriptors, making them poorly suited for structured data, such as molecular data(Sun et al., 2020).

Over recent years, graph-based DL has garnered considerable attention in computational drug development owing to its robust capability to capture complex and interrelated relationships between various biological entities in a structured format (Partin et al., 2023). An increasing number of studies have adopted graph-based methods for DRP. For instance, Nguyen et al. (G. T. T. Nguyen et al., 2022) extracted drug features from their graph representation using graph convolutional networks (GCN) and incorporated multi-omics data with convolutional neural networks (CNN) for predicting response values. Peng et al. (Peng, Chen, et al., 2022) designed a heterogeneous network that integrates cell line-drug similarities and associations, performed graph convolution to learn latent embeddings. Liu et al. (X. Liu et al., 2022) predicted cancer drug response by constructing a graph neural network (GNN) using multi-omics data, drug chemical structures and known responses while incorporating contrastive learning for enhancing generalization.

Despite the significant progress made by graph-based approaches in DRP, there are still three major issues that need to be addressed. (1) Existing graph-based methods primarily focus on pairwise correlations among nodes, leading to the aggregation of information only from neighboring nodes, while failing to capture high-order relations that necessitate considering intricate interactions among cell lines (Jin et al., 2023b). (2) Most current approaches for extracting drug features involve converting drugs' SMILES strings into graph structures and simply utilizing GNN to learn drug representation. However, the structure of a drug molecule contains diverse types of structural information, such as functional groups and chemical bonds. The significance and interactions of these different types of information have not been sufficiently considered during the modeling process (Chu et al., 2023; Hu et al., 2024). (3) Previous research has not fully utilized cell line-drug association, which can potentially provide valuable auxiliary support for DRP. Although several recent studies have attempted to address this limitation, they typically focus on modeling the interrelations between drugs and cell lines based on sensitive or resistant responses, neglecting the complementary information offered by both types of responses (X. Liu et al., 2022; Xu et al., 2024).

To tackle the above issues, we propose HCLGT-DRP, a novel computational framework based on hypergraph contrastive learning and graph transformer for DRP. Specifically, we first construct a hypergraph for each type of omics data and introduce a hypergraph aggregation mechanism to optimize structural information. To integrate multi-omics data, a hypergraph fusion method is proposed for capturing high-order relations across multiple omics. Then, we incorporate a cell line-drug association graph to provide supplementary information, while applying contrastive learning to obtain representations of cell lines. In addition, a graph transformer with molecular-specific attention is designed, which simultaneously considers both drug structure and feature information to learn drugs representations. Lastly, the cell line-drug response matrix is reconstructed by integrating cell lines and drugs representations using the linear correlation coefficient.

The main contributions can be summarized as follows:

- We introduce a hypergraph aggregation mechanism to more accurately represent the high-order relations within the hypergraph structure of various types of omics data.
- We propose a hypergraph fusion method to integrate omics-specific hypergraphs into a unified one, enhancing the representation of complex multi-relational data across multiple omics.
- We design a graph transformer with molecular-specific attention to facilitate the extraction of more effective drug representations, improving graph representation learning ability.
- Extensive experiments on two public datasets show that HCLGT-DRP outperforms ten representative baselines across various DRP tasks.

2. Related work

2.1. Graph-based drug response prediction

Existing graph-based methods for DRP are broadly divided into two primary categories based on differing graph construction strategies.

The first category involves abstracting cell lines or drugs into graph representations (G. T. T. Nguyen et al., 2022; Xu et al., 2024). Nguyen et al. (T. Nguyen et al., 2022) represented drugs as molecular graphs, where atoms are treated as nodes and chemical bonds as edges, with their features representing the physicochemical properties of the atoms and the molecular structure. Additionally, cell lines were encoded as binary vectors representing genomic aberrations, learning their representations through convolutional layers and combining them for DRP. Liu et al. (Q. Liu et al., 2020) proposed a hybrid GCN that combines a uniform GCN with multiple subnetworks to automatically learn the latent representation of drug structures and assess their responses. However, these methods primarily focus on the representations of individual drugs or cell lines while neglecting their crucial interactions.

The second category concentrates on modeling the interactions between cell lines and drugs, with nodes representing drugs and cell lines, and edges denoting known responses (Peng, Chen, et al., 2022; Tao et al., 2023). Peng et al. (Peng, Liu, et al., 2022) used two GCNs to simultaneously learn features of cell lines and drugs, incorporating interactions between nodes and their neighbors to enhance feature learning and improve DRP. A recently proposed model, named TSGCNN (Peng et al., 2023), introduced drug and cell line similarities in homogeneous nodes and employed a dual-space GCN to learn features from both homogeneous and heterogeneous neighbors, eventually predicting drug responses through the reconstruction of drug-cell line association.

Graph transformer provides a novel neural network for graph-structured data (Shehzad et al., 2024). Unlike GCNs and graph attention networks (GAT) (Veličković et al., 2018), which employ fixed convolution and attention heads, graph transformer adapts attention to prioritize relevant node connections (J. Liu et al., 2024), and has increasingly been applied in the biomedical field. Chu et al. (Chu et al., 2023) combined graph transformer and GAT-GCN to enhance drug representation, integrated multi-omics data using CNN, and reduced noise in transcriptomics data with KernelPCA before predicting IC50 values. Despite these advancements, existing graph-based models face challenges in entirely using the structural information of drug data and adequately addressing higher-order relationships between drugs and cell lines.

2.2. Hypergraph learning

Hypergraphs facilitate modeling relationships encompassing more than two entities simultaneously, unlike traditional graph structures that primarily focus on pairwise relations (Jiang et al., 2019). In a hypergraph, an edge (a hyperedge) can connect multiple nodes, making it an ideal framework for representing multi-way interactions (Zhou et al., 2023). Hypergraph learning has demonstrated tremendous potential across various biomedical applications (Saifuddin et al., 2023; Xie et al., 2024). For instance, Wang et al. (H. Wang et al., 2024) introduced HyperTMO, a multi-omics integration approach using hypergraph convolutional networks (HGCN), which employs hypergraph structures and evidence theory for patient classification. Peng et al. (Peng et al., 2024) proposed HRLCDR, a computational method leveraging hypergraph representation learning to predict cancer drug response, which captures higher-order interactions through heterogeneous graph convolutions and hypergraph convolutions. However, the application of hypergraphs in DRP remains limited and their potential has yet to be fully explored and validated.

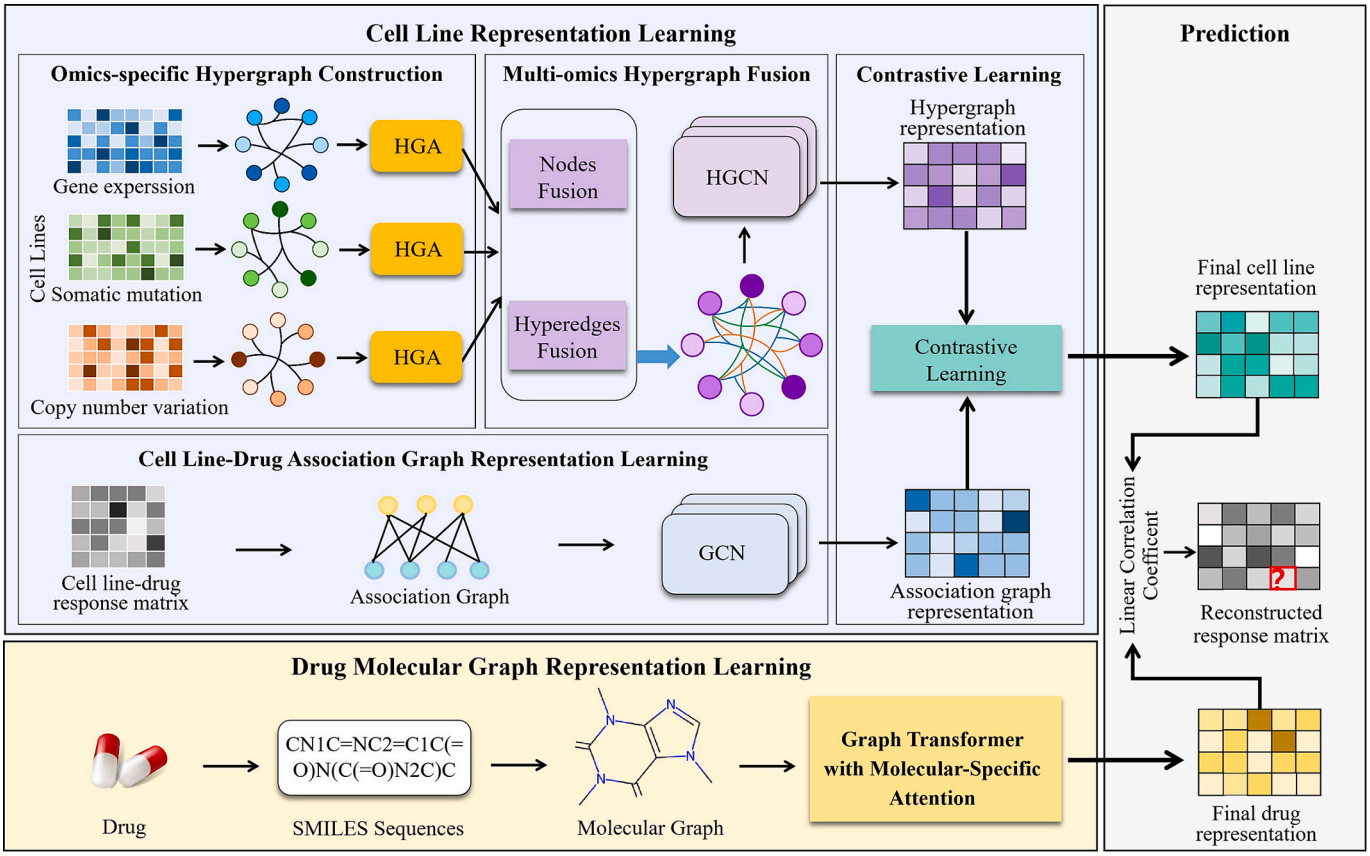


Fig. 1. The workflow of HCLGT-DRP for drug response prediction.

3. Methodology

3.1. Problem definition

In this study, drug response prediction is formulated as a supervised classification task. Given m cell lines and n drugs, the prediction task is to forecast the binary drug response outcome $R \in \{0,1\}^{m\times n}$. Cell lines are characterized by multi-omics profiles, including gene expression (gne), somatic mutation (mu), and copy number variation (cnv). Drugs are represented as canonical SMILES format. In the 2D response matrix $R \in \{0,1\}^{m\times n}$, every entry $r_{ij} \in R^{m\times n}$ represents the response of the *i-th* cell line to the *j-th* drug, where $r_{ij} = 1$ denotes the *i-th* cell line is sensitive to the *j-th* drug and $r_{ij} = 0$ denotes the *i-th* cell line is resistant or unknown to the *j-th* drug.

3.2. Overview of HCLGT-DRP

The overall framework of HCLGT-DRP is depicted in Fig. 1, addressing DRP in three steps.

(1) Cell line representation learning incorporates hypergraph representation learning for multi-omics profiles of cell lines along with graph representation learning for cell line-drug association to capture high-order relationships among various omics data, obtaining high-quality representations of cell lines.

(2) Drug representation learning employs a graph transformer with molecular-specific attention that explores both the drug molecular graph structures and atomic distances to learn more effective drug representations.

(3) Drug response prediction calculates the linear correlation coefficients between the ultimate representations of cell lines and drugs to reconstruct the cell line-drug response matrix, predicting whether cell lines exhibit sensitivity or resistance to drugs.

3.3. Cell line representation learning

3.3.1. Hypergraph construction for omics data

Based on the idea that similar cell lines tend to exhibit similar drug sensitivity (N. Zhang et al., 2015), we construct a hypergraph for single-omics data by utilizing cell line similarity and known drug response to explore high-order relationships of cell lines. The following formula calculates the single-omics similarity firstly:

$$S_{ij} = \begin{cases} e^{-\frac{\|X_i - X_j\|_2}{2\varepsilon^2}} & j \in T_i \\ 0 & j \notin T_i \end{cases} \tag{1}$$

where X_i is the Gaussian-regularized omics expression values of *i*-th cell line. T_i is the set of top T cell lines that are most similar to *i*-th cell line, with T fixed at 10 determined through grid search. The scale parameter ε is set to 3 followed by (Peng, Chen, et al., 2022; Peng et al., 2024). Inspired by (Jin et al., 2023a), we obtain high-order similarity matrix A within single omics:

$$A_1 = (S \cdot S) \odot S \tag{2}$$

$$A_2 = \left(R \cdot R^T\right) \odot S \tag{3}$$

$$A_3 = \left(S \cdot R^T\right) \cdot \left(S \cdot R^T\right)^T \tag{4}$$

$$A = A_1 + A_2 + A_3 \tag{5}$$

where R is the drug response matrix. $\odot$ is the element-wise product. A_1 captures the similarity between cell lines, A_2 reflects the similarity in drug responses among similar cell lines, and A_3 represents an indirect correlation of similar cell lines. Thus, A captures the higher-order similarities between cell lines. Then, we construct a *k-uniform* hypergraph

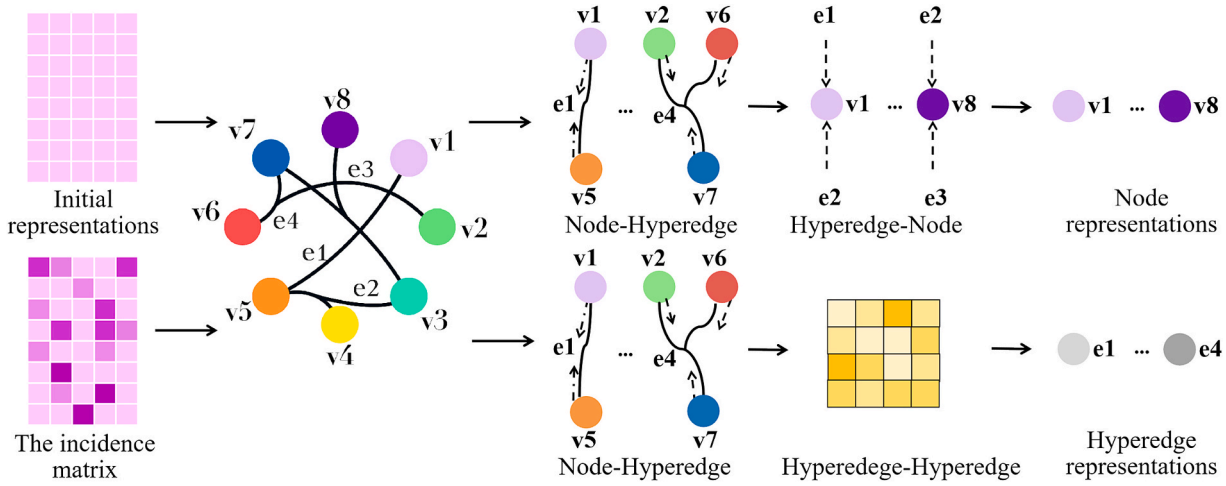


Fig. 2. Illustration of the HGA mechanism with $k = 2$.

using the K-Nearest Neighbors (KNN) algorithm. Specifically, each cell line is represented as a vertex. For each vertex, its top-K nearest neighbors are grouped into a hyperedge based on high-order similarity. K is set to 10. The structure of a hypergraph can be defined by its incidence matrix H, which can be denoted as follows:

$$H(v,e) = \begin{cases} 1 & if\ \ v \in e \\ 0 & otherwise \end{cases} \tag{6}$$

The vertex degree matrix D_v is a diagonal matrix, which defined as follows:

$$D_v(i,i) = \sum_{e \in E} H(i,e) \tag{7}$$

Similarly, the hyperedge degree matrix D_e is defined as follow:

$$D_e(i,i) = \sum_{v \in V} H(v,i) \tag{8}$$

The raw omics data is processed through 1D convolution to serve as the initial node features X.

As the contribution of different nodes and hyperedges to the final representations varies, we introduce the Hypergraph Aggregation (HGA) mechanism to optimize the hypergraph structure within each omics. As shown in Fig. 2, given $H = (E = \{e_1, e_2, \ldots\}, \mathscr{V} = \{v_1, v_2, \ldots\})$, the representation of the hyperedge e_i can be formulated as follows:

$$e_i = avg\left(\sum_{v_j \in N(e_i)} v_j\right) \tag{9}$$

where avg is the averaging operator $avg(x) = \frac{1}{M}\sum_{i=1}^{M} x_i$ and M refers to the number of neighboring hyperedges of node v_i. $N(e_i)$ is the neighboring nodes connected by e_i. We employ a hyperedge attention for hyperedge representations $E = \{e_1, e_2, e_3, \ldots.\}$. Three separate linear projections are performed to the hyperedge representations to get queries(Q), key(K), Value(V) by the hyperedge representations:

$$Y = softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{D_K}}\right)V \tag{10}$$

where D_k is the dimension of Q and K vectors. Y is the updated hyperedge representations.

The interaction between each node $v_i \in \mathscr{V}$ and the relevant hyperedges is captured, which can be expressed as follows:

$$H'_{ij} = \frac{\exp\left(LeakyReLU\left(MLP\left(Concat\left[v_i, e_j\right]\right)\right)\right)}{\sum_{e_s \in N(v_i)} \exp\left(LeakyReLU\left(MLP\left(Concat\left[v_i, e_s\right]\right)\right)\right)} \tag{11}$$

where $Concat\left[v_i, e_j\right]$ is the concatenation of the representation vector of v_i and its associated hyperedges e_j. $N(v_i)$ is the neighboring hyperedges connected to v_i. $MLP(\cdot)$ refers to the multi-layer perceptron. $LeakyReLU(\cdot)$ is the activation function. H'_{ij} denotes the attention weight between node v_i and hyperedge e_j. Leveraging the normalized Laplacian operator, we apply hypergraph convolution operation as follows:

$$X^{(k)} = LeakyReLU\left(D_v^{-1/2}\left(H'\right)D_e^{-1}\left(H'\right)^{\mathrm{T}}D_v^{-1/2}X^{(k-1)}\theta\right) \tag{12}$$

where $X^{(k)}$ is the updated node representations at the k-th layer and $X^{(0)} = X$. k controls propagation steps. D_e and D_v are diagonal matrices representing the degrees of hyperedges and nodes, respectively. H' is the weighted incidence matrix and θ is the learnable weight matrix.

3.3.2. Multi-omics hypergraph fusion

To explore the relationships across different omics data, we fuse multiple omics-specific hypergraphs into a unified one. Inspired by multi-view learning (Y. Liu et al., 2023), we recognize that the quality of cell line representations varies across omics. A graph-level attention mechanism is introduced to integrate the significance of single-omics hypergraphs.

Our method incorporates three omics types: gene expressions, somatic mutations and copy number variations. First, let $X' = Concat\left[H'_{gne}, H'_{mu}, H'_{cnv}\right]$ be the concatenation of nodes' representations after applying the HGA within three omics. Then, we use a set of parameter vectors $z = \left\{z_{\mathrm{gne}}, z_{mu}, z_{cnv}\right\}$ to generate the omics weights.

$$\alpha'_i = avg(X' z_i + b) \tag{13}$$

where $i \in \{gne, mu, cnv\}$ and b is the bias. The graph-level attention coefficient is normalized as follows:

$$\alpha_i = \frac{\exp\left(\alpha'_i\right)}{\exp\left(\alpha'_{gne}\right) + \exp\left(\alpha'_{mu}\right) + \exp\left(\alpha'_{cnv}\right)} \tag{14}$$

The graph-level attention weight can be denoted as $\alpha = \left[\alpha_{gne}, \alpha_{mu}, \alpha_{cnv}\right]$. Hence, the fused node representation across three omics is calculated through a weighted combination as follows:

$$X^* = READOUT\left(MLP\left(\alpha_{gne} \cdot H'_{gne} + \alpha_{mu} \cdot H'_{mu} + \alpha_{cnv} \cdot H'_{cnv}\right)\right) \tag{15}$$

where READOUT is specified as concatenation by row vectors.

During the hyperedges fusion, we focus on the potential similarities between hyperedges from different omics. Specifically, some sub-

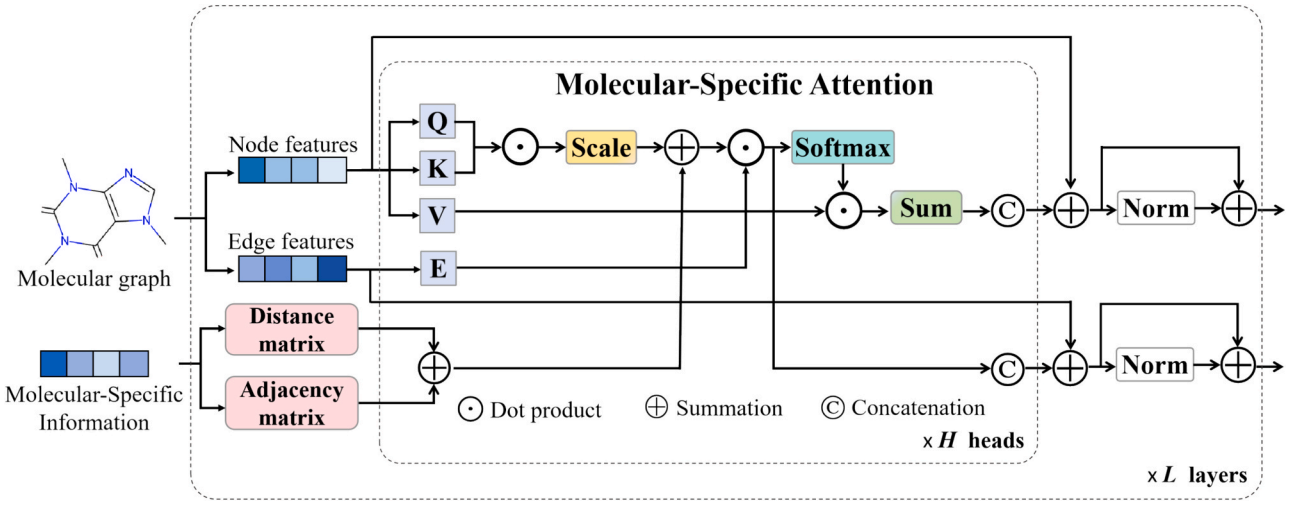


Fig. 3. The network structure of a graph transformer with molecular-specific attention.

relations can appear in multiple omics and result in the same hyperedge connections, referring to shared hyperedges.

$$E_{shared} = E_{gne} \cap E_{mu} \cap E_{cnv} \tag{16}$$

where E_{shared} denotes the set of hyperedges that are commonly shared across the three single-omics hypergraphs. We apply the graph-level attention mechanism for these shared hyperedges to fuse cross-omics hyperedges representations.

$$Y^{*}_{shared} = READOUT\left(MLP\left(\alpha_{gne} \cdot E'_{gne} + \alpha_{mu} \cdot E'_{mu} + \alpha_{cnv} \cdot E'_{cnv}\right)\right) \tag{17}$$

where READOUT is specified as concatenation by row vectors.

Beyond the shared hyperedges, non-shared hyperedges represent unique connection patterns specific to individual hypergraphs. These unique hyperedges can reveal distinct interactions within particular omics.

$$E_{unique} = \left(E_{gne} \cup E_{mu} \cup E_{cnv}\right) \backslash E_{shared} \tag{18}$$

where E_{unique} denotes the set of hyperedges that are exclusive to a single omics-specific hypergraph and are not shared across the three omics. The representations of the unique hyperedges Y^{*}_{unique} are preserved as computed in their respective original omics-specific hypergraphs. Thereby, we obtain all hyperedges $E_{fused} = E_{shared} \cup E_{unique}$ and their representations $Y_{fused} = Concat\left[Y^{*}_{shared}, Y^{*}_{unique}\right]$ in the fused hypergraph. The fused hypergraph is first processed through HGA to obtain the final hypergraph weighted incidence matrix H^{*}. Then a hypergraph convolutional layer is employed to learn the final representations of cell lines as follows:

$$X_{final} = softmax(D_v^{-1/2}(H^{*})D_e^{-1}(H^{*})^{T}D_v^{-1/2}X^{*}\Theta^{*}) \tag{19}$$

where Θ^{*} is the parameter matrix. X_{final} represents the final cell lines representations through hypergraph learning.

3.3.3. Cell line-drug association graph learning

Previous studies have overlooked the valuable information embedded in known drug responses, which can be utilized to construct a response matrix based on their sensitivity and resistance as follows:

$$R_{ij} = \begin{cases} 1, & c_i \text{ is sensitive to } d_i \\ 0, & c_i \text{ is resistant to } d_i \end{cases} \tag{20}$$

where $R_{ij} \in R^{m \times n}$, with m denoting the number of cell lines and n denoting the number of drugs. Specifically, the known cell line-drug association is represented as a heterogeneous graph, where nodes represent drugs and cell lines, and edges indicate known associations. The node representations, symbolized as $X = \begin{bmatrix} X^{C} \\ X^{D} \end{bmatrix}$, are trainable vectors initialized randomly, where X^{C} corresponds to the representations of cell lines and X^{D} represents the representations of drugs. We integrate both sensitive and resistant response information to a graph structure, adjusting edge weights to consider both types of responses. Sensitive edges connect drugs and cell lines with high sensitivity, having a higher initial weight, while resistant edges have a lower initial weight. The weight matrix A can be described as follows:

$$A_{ij} = \begin{cases} \alpha R_{ij} & \text{if } R_{ij} = 1 \\ \beta(1 - R_{ij}) & \text{if } R_{ij} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where α and β are the weight for sensitive and resistant edges, respectively, and satisfy $\alpha > \beta$. The representation of each cell line node is updated iteratively through graph convolution layers. The $(l + 1)$-th layer output of the association graph can be represented as follows:

$$H^{(l+1)} = LeakyReLU\left(D^{-1/2}AD^{-1/2}H^{(l)}W^{(l)}\right) \tag{22}$$

where D is a diagonal degree matrix of the graph, and $H^{(0)} = X$. The representations of all cell line learned from the association graph can be defined as $H = Concat[h_0, \ldots, h_m]$, where $h_i = Concat\left[h_i^{(0)}, h_i^{(1)}, \cdots, h_i^{(L)}\right]$ is the concatenation of the outputs from all layers, and L is the number of layers of GCN.

3.3.4. Contrastive learning

To preserve both the cell line features derived from hypergraph learning and association graph learning, we adopt a hypergraph contrastive learning approach. It effectively integrates intrinsic cell-specific feature with extrinsic drug response behavior. This method is designed to capture intra-node consistency as well as inter-node dissimilarities across the different learning frameworks. Specifically, for the *i-th* cell line, two representations are generated: one from hypergraph learning process denoted as u_i and another from association graph learning process denoted as v_i. These two representations form a positive pair and representations corresponding to different cell lines are treated as negative pairs. Negative samples are constructed in two ways: (i) intra-negative pairs (u_k with $k \neq i$) are formed when the representation of other nodes learned by the same learning process as the anchor u_i and (ii) inter-negative pairs (v_k with $k \neq i$) are formed when inter-negative pairs is not learned by the same learning process as the anchor u_i. We

apply InfoNCE (Oord et al., 2019) to enhance positive pair similarities while suppressing negative pair similarities. For m cell lines, the contrastive loss of the hypergraph learning is expressed as follows:

$$\mathscr{L}_{HGF}(u_i, v_i) = -\sum_{i=1}^{m} \log \frac{\exp(sim(u_i, v_i)/\tau)}{\exp(sim(u_i, v_i)/\tau) + \sum_{k\neq i}\exp(sim(u_i, u_k)/\tau) + \sum_{k\neq i}\exp(sim(u_i, v_k)/\tau)} \tag{23}$$

where $sim(\cdot)$ is the similarity function, with cosine similarity used in this study. τ is temperature parameter, and it is set to 0.5.

Since the two views learned by hypergraph learning and association graph learning for m cell lines are symmetric, the loss of the association graph learning can be defined as follows:

$$\mathscr{L}_{AGF}(v_i, u_i) = -\sum_{i=1}^{m} \log \frac{\exp(sim(v_i, u_i)/\tau)}{\exp(sim(v_i, u_i)/\tau) + \sum_{k\neq i}\exp(sim(v_i, v_k)/\tau) + \sum_{k\neq i}\exp(sim(v_i, u_k)/\tau)} \tag{24}$$

The overall contrastive loss function is as follows:

$$\mathscr{L}_{cl} = \lambda\mathscr{L}_{HGF} + (1-\lambda)\mathscr{L}_{AGF} \tag{25}$$

where λ is a hyper-parameter that balances the contributions of different loss functions.

3.4. Drug molecular representation learning

The original drug SMILES string is converted into a molecular graph. Rather than employing a transformer to represent chemical molecules as sentences, an enhanced graph transformer is proposed, which incorporates the structural information of the molecules. This approach eliminates the requirement for linear representations of molecules as inputs. The self-attention mechanism is improved by integrating the molecular graph structure and atomic distance, as depicted in Fig. 3.

Let A denote the adjacency matrix of the drug molecular graph, where A_{ij} indicates the presence or absence of an edge (bond) between nodes i and j. Let $Dist$ represent the inter-atomic distance matrix, where $Dist_{ij}$ is the relative distance between atoms i and i, and $Dist$ is obtained using the RDKit package (Bento et al., 2020). The detailed process for updating representations in the l-th layer is as follows:

$$Q_{ij}^{k,l} = Q^{k,l}Norm\left(h_i^l\right), K_{ij}^{k,l} = K^{k,l}Norm\left(h_j^l\right), V_{ij}^{k,l} = V^{k,l}Norm\left(h_j^l\right) \tag{26}$$

$$E_{ij}^{k,l} = E^{k,l}Norm\left(e_{ij}^l\right) \tag{27}$$

$$w_{ij}^{k,l} = softmax_j\left(\left(\frac{Q_{ij}^{k,l}h_i^l \cdot K_{ij}^{k,l}h_j^l}{\sqrt{D_k}} + Dist_{ij} + A_{ij}\right)\cdot E_{ij}^{k,l}e_{ij}^l\right) \tag{28}$$

$$\widetilde{h}_i^{l+1} = h_i^l + O_h^l\left(Concat_{k=1}^{H}\left(\sum_{j\in N_i} w_{ij}^{k,l}V_{ij}^{k,l}h_j^l\right)\right) \tag{29}$$

$$\widetilde{e}_{ij}^{l+1} = e_{ij}^l + O_e^l\left(Concat_{k=1}^{H}\left(w_{ij}^{k,l}\right)\right) \tag{30}$$

where $Q^{k,l}, K^{k,l}, V^{k,l}, E^{k,l}, O_h^l, O_e^l$ are the parameters of the linear layers. The quantity of attention heads is indicated by H. D_k is the dimension of the query and key vectors. The operations of batch normalization and concatenation are represented by $Norm$ and $Concat$, respectively. $\widetilde{h}_i^{l+1}$ and $\widetilde{e}_{ij}^{l+1}$ are passed to feed forward networks that incorporate residue connections and normalization layers to learn the final output nodes representations of the $(l+1)$-th layer.

Table 1
A summary of the statistics for three datasets.

Dataset	Drug	Cell Line	Sensitive Sample	Resistant Sample
GDSC	228	962	20,851	156,512
CCLE	24	436	1,696	8,768
TCGA	22	403	201	229

3.5. Drug response prediction

The linear correlation coefficient is utilized as a decoder to reconstruct the drug response matrix after obtaining the representations of cell lines and drugs.

$$Corr\left(h_i^c, h_j^d\right) = \frac{\left(h_i^c - \mu_i^c\right)\left(h_j^d - \mu_j^d\right)^T}{\sqrt{\left(h_i^c - \mu_i^c\right)\left(h_i^c - \mu_i^c\right)^T}\sqrt{\left(h_j^d - \mu_j^d\right)\left(h_j^d - \mu_j^d\right)^T}} \tag{31}$$

where h_i^c and h_j^d are the final representations of the i-th cell line and the j-th drug, respectively. μ_i^c is the means of h_i^c and μ_j^d is the means of h_j^d. The correlation coefficient ranges from $[-1,1]$. Then, the following function is employed to activate the output:

$$\varphi(h) = \frac{e^{-\alpha h}}{1+e^{-\alpha h}} \tag{32}$$

The reconstruction of the association matrix R^* is performed by

$$R^* = \varphi\left(Corr\left(h^c, h^d\right)\right) \tag{33}$$

The loss function for the classification task is formulated as follows:

$$\mathscr{L}(R, R^*) = -\frac{1}{m\times n}\sum_{i,j} M_{ij}\left[R_{ij}log\left(R_{ij}^*\right) + \left(1-R_{ij}\right)log\left(1-R_{ij}^*\right)\right] \tag{34}$$

where R and R^* are the response matrix before and after reconstruction, respectively. M is a binary indicator matrix, where $M_{ij} = 1$ if the response between the i-th cell line and the j-th drug exists, otherwise $M_{ij} = 0$.

4. Experimental settings

4.1. Datasets

To assess the performance of HCLGT-DRP, comprehensive experiments are conducted on two public datasets, namely GDSC and CCLE. We use the preprocessed data provided by MOFGCN(Peng, Chen, et al., 2022), which follows the preprocessing strategy described in MOLI (Sharifi-Noghabi et al., 2019). The cell line-drug response matrix is converted into a binary matrix using the thresholds provided by the GDSC dataset. For the CCLE dataset, responses are classified based on activity area values, with sensitivity for Z-score normalized values above 0.8, and resistance otherwise (Dong et al., 2015). To further assess the model's ability to generalize from *in vitro* to *in vivo* data, an additional external validation is conducted using the Cancer Genome Atlas (TCGA) dataset. As in previous studies (Ding et al., 2016), clinical annotations of patient drug treatments were obtained. Annotations with "Complete Response" and "Partial Response" are categorized as sensitive samples (Peng, Liu, et al., 2022). Omics data were processed as (Sharifi-Noghabi et al., 2019). Table 1 summarizes the statistical information for these datasets.

In addition, we download SMILES strings for drugs from the PubChem (Kim et al., 2019). Using the complexation methods in the DeepChem library (Duvenaud et al., 2015), each SMILES string is

converted into a molecular graph. The attributes of each atomic node are encoded as a 75-dimensional vector.

4.2. Baseline methods

We compare HCLGT-DRP with ten state-of-the-art baselines to verify its effectiveness.

- **SRMF** (L. Wang et al., 2017): A similarity-regularized matrix factorization approach for predicting drug responses by drug chemical structures and gene expressions.
- **MOLI** (Sharifi-Noghabi et al., 2019): A deep learning network-based multi-omics late integration approach for DRP.
- **DeepForest** (Su et al., 2019): A deep cascaded forest method for classifying drug responses.
- **DeepCDR** (Q. Liu et al., 2020): A hybrid GCN model that predicts cancer drug response by integrating multi-omics data and drug chemical structures.
- **GraOmicDRP** (G. T. T. Nguyen et al., 2022): A DL-based method to integrate molecular graphs of drugs and multi-omics data for DRP.
- **MOFGCN** (Peng, Chen, et al., 2022): A multi-omics fusion method for DRP that utilizes GCN to learn a heterogeneous network.
- **GraphDRP** (T. Nguyen et al., 2022): A GCN-based method that predicts drug-cell line response by learning representative features from molecular graphs of drugs and binary genomic aberration vectors of cell lines.
- **GraphCDR** (X. Liu et al., 2022): A GNN-based method that integrates multi-omics with contrastive learning to predict cancer drug responses.
- **DGCL** (Tao et al., 2023): A dynamic hypergraph contrastive learning method that leverages local and global interrelations between genes and drugs to predict drug-gene interaction.
- **RedCDR** (Xu et al., 2024): A dual relation distillation model that enables independent learning and interactive fusion of cell line-drug attributes and cell line-drug association.

4.3. Experimental design

To comprehensively assess the proposed HCLGT-DRP from various aspects, we design the following experiments:

- Test 1: Random zeroing test. Randomly removing several known cell line-drug associations for predicting the missing drug responses.
- Test 2: A new cell line or drug prediction. Blinding a row or column in the response matrix randomly to predict responses for an unseen cell line or drug.
- Test 3: A single drug prediction. Randomly selecting a drug to predict responses across various cell lines.
- Test 4: Ablation studies. Ablating different components of HCLGT-DRP to evaluate their performance contributions.
- Test 5: Visualization analysis. Performing visualization of representations to assess the learning ability of HCLGT-DRP.
- Test 6: Hyper-parameters sensitivity analysis. Investigating the impact of important hyper-parameters of HCLGT-DRP on its performance.
- Test 7: External validation on *in vivo* data. Training our model on *in vitro* data and evaluating its performance on *in vivo* drug response data.
- Test 8: Case study on unknown responses. Conducting a case study to validate the predictive capability of HCLGT-DRP for novel cell line-drug responses.

4.4. Evaluation metrics

The area under the receiver operating characteristic curve (AUC), accuracy (ACC), precision, and F1 Score (F1) are applied to assess the compared methods. The results are reported as the mean and standard deviation (mean ± std) of these metrics across experiments using different random seeds. To ensure a fair comparison, all compared models employ identical multi-omics and drug data as inputs while adhering to hyperparameter configurations defined in their original literature. All experiments are conducted on a machine equipped with a GeForce RTX 4070 GPU (16 GB memory) and an Intel(R) Core(TM) i7-13700 K CPU @ 5.40 GHz. All algorithms are implemented in PyTorch and executed using Python 3.8.16.

Table 2
Performance of different methods under randomly zeroing cross-validation on two datasets. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Model	AUC	ACC	Precision	F1 Score
GDSC	SRMF	$0.6572 \pm 4 \times 10^{-4}$	$0.5691 \pm 2 \times 10^{-4}$	$0.5309 \pm 4 \times 10^{-4}$	$0.6718 \pm 2 \times 10^{-4}$
	MOLI	$0.7244 \pm 1 \times 10^{-4}$	$0.6547 \pm 4 \times 10^{-4}$	$0.6219 \pm 3 \times 10^{-4}$	$0.6981 \pm 7 \times 10^{-5}$
	DeepForest	$0.7292 \pm 3 \times 10^{-4}$	$0.6503 \pm 1 \times 10^{-4}$	$0.6217 \pm 2 \times 10^{-4}$	$0.6894 \pm 4 \times 10^{-5}$
	DeepCDR	$0.7649 \pm 4 \times 10^{-5}$	$0.7061 \pm 5 \times 10^{-4}$	$0.6896 \pm 6 \times 10^{-4}$	$0.7307 \pm 2 \times 10^{-5}$
	GraOmicDRP	$0.7531 \pm 1 \times 10^{-5}$	$0.6609 \pm 2 \times 10^{-5}$	$0.6394 \pm 4 \times 10^{-5}$	$0.7104 \pm 4 \times 10^{-5}$
	MOFGCN	$0.8665 \pm 3 \times 10^{-6}$	$0.7768 \pm 1 \times 10^{-5}$	$0.7473 \pm 4 \times 10^{-5}$	$0.7902 \pm 2 \times 10^{-5}$
	GraphDRP	$0.7764 \pm 2 \times 10^{-5}$	$0.6829 \pm 4 \times 10^{-5}$	$0.6513 \pm 2 \times 10^{-5}$	$0.7609 \pm 3 \times 10^{-5}$
	GraphCDR	$0.8143 \pm 6 \times 10^{-4}$	$0.7088 \pm 2 \times 10^{-3}$	$0.6813 \pm 1 \times 10^{-3}$	$0.7686 \pm 5 \times 10^{-3}$
	DGCL	$0.7935 \pm 3 \times 10^{-6}$	$0.7413 \pm 4 \times 10^{-6}$	$0.7393 \pm 1 \times 10^{-5}$	$0.7716 \pm 3 \times 10^{-6}$
	RedCDR	$\underline{0.8692 \pm 1 \times 10^{-5}}$	$\underline{0.7804 \pm 6 \times 10^{-5}}$	$\underline{0.7660 \pm 4 \times 10^{-5}}$	$\underline{0.7994 \pm 6 \times 10^{-5}}$
	HCLGT-DRP	$\mathbf{0.8796 \pm 4 \times 10^{-5}}$	$\mathbf{0.8034 \pm 2 \times 10^{-5}}$	$\mathbf{0.7961 \pm 3 \times 10^{-5}}$	$\mathbf{0.8317 \pm 4 \times 10^{-5}}$
CCLE	SRMF	$0.7661 \pm 5 \times 10^{-5}$	$0.6797 \pm 3 \times 10^{-5}$	$0.6213 \pm 5 \times 10^{-5}$	$0.7432 \pm 1 \times 10^{-5}$
	MOLI	$0.7863 \pm 1 \times 10^{-4}$	$0.7059 \pm 5 \times 10^{-4}$	$0.6953 \pm 3 \times 10^{-4}$	$0.7459 \pm 3 \times 10^{-4}$
	DeepForest	$0.7106 \pm 2 \times 10^{-4}$	$0.6581 \pm 3 \times 10^{-4}$	$0.6414 \pm 1 \times 10^{-4}$	$0.6900 \pm 2 \times 10^{-4}$
	DeepCDR	$0.8029 \pm 2 \times 10^{-4}$	$0.7667 \pm 1 \times 10^{-4}$	$0.7590 \pm 3 \times 10^{-4}$	$0.7712 \pm 3 \times 10^{-4}$
	GraOmicDRP	$0.7280 \pm 5 \times 10^{-5}$	$0.6741 \pm 5 \times 10^{-5}$	$0.6248 \pm 4 \times 10^{-5}$	$0.6740 \pm 3 \times 10^{-5}$
	MOFGCN	$\underline{0.8603 \pm 2 \times 10^{-4}}$	$0.7614 \pm 2 \times 10^{-4}$	$0.7695 \pm 5 \times 10^{-4}$	$0.7936 \pm 1 \times 10^{-4}$
	GraphDRP	$0.7821 \pm 6 \times 10^{-4}$	$0.7048 \pm 1 \times 10^{-5}$	$0.6817 \pm 2 \times 10^{-5}$	$0.7506 \pm 2 \times 10^{-5}$
	GraphCDR	$0.8442 \pm 3 \times 10^{-5}$	$\underline{0.7762 \pm 2 \times 10^{-5}}$	$0.7705 \pm 2 \times 10^{-5}$	$0.7894 \pm 1 \times 10^{-5}$
	DGCL	$0.7897 \pm 2 \times 10^{-6}$	$0.7517 \pm 4 \times 10^{-6}$	$0.7501 \pm 5 \times 10^{-5}$	$0.7769 \pm 1 \times 10^{-6}$
	RedCDR	$0.8453 \pm 2 \times 10^{-5}$	$0.7761 \pm 4 \times 10^{-5}$	$\underline{0.7747 \pm 3 \times 10^{-5}}$	$\underline{0.7958 \pm 3 \times 10^{-5}}$
	HCLGT-DRP	$\mathbf{0.8703 \pm 1 \times 10^{-5}}$	$\mathbf{0.8046 \pm 3 \times 10^{-5}}$	$\mathbf{0.7854 \pm 3 \times 10^{-5}}$	$\mathbf{0.8103 \pm 2 \times 10^{-5}}$

5. Results and discussions

5.1. Randomly zeroing cross-validation (Test 1)

To investigate the effectiveness of HCLGT-DRP to predict missing drug responses, we randomly remove a specified proportion of known responses and compare its performance with all baselines. We apply a five-fold cross-validation, repeated five times. Specifically, the known cell line-drug responses are randomly divided into five equal subsets. In each iteration, one subset is designated as the testing set, while the other four are used as training set.

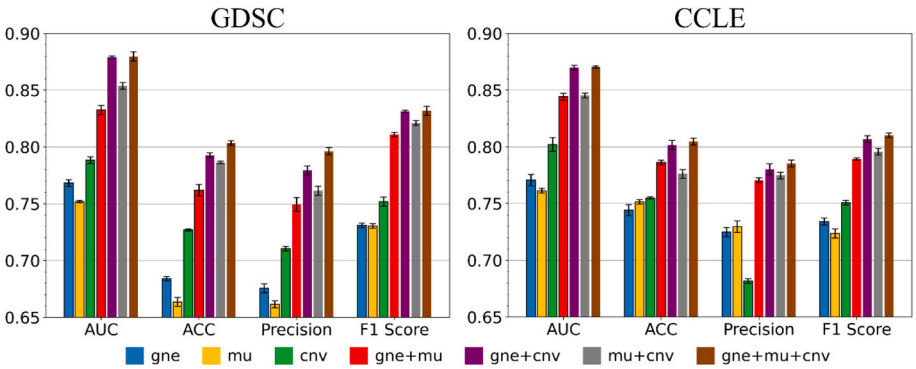


Fig. 4. Performance comparisons under different combinations of omics data types on the GDSC and CCLE datasets.

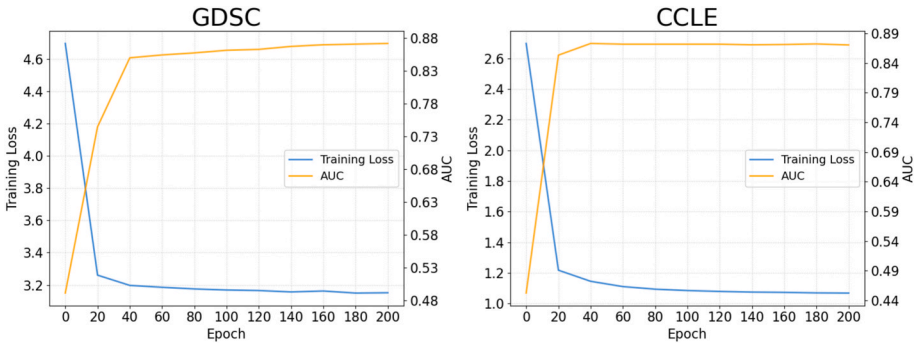


Fig. 5. Convergence analysis of HCLGT-DRP on the GDSC and CCLE datasets.

Table 2 indicates that HCLGT-DRP consistently outperforms all ten baseline methods across both datasets, achieving superior results on all metrics and demonstrating its strong potential in predicting missing drug responses. Specifically, on the GDSC dataset, HCLGT-DRP surpasses the best-performing baseline, RedCDR, with improvements of 1.04 %, 2.30 %, 3.01 %, and 3.23 % in AUC, ACC, precision, and F1, respectively. Similarly, on the CCLE dataset, HCLGT-DRP achieves higher performance compared to the suboptimal methods, with a 1.00 % increase in AUC, a 2.85 % elevation in ACC, a 1.07 % boost in precision, and a 1.45 % improvement in F1.

Notably, the graph-based approaches (DeepCDR, GraOmicDRP, MOFGCN, GraphDRP, GraphCDR, DGCL, RedCDR and HCLGT-DRP) tend to outperform DL methods without graph structures (SRMF, MOLI and DeepForest). This performance is mainly driven by the high-quality representations of cell lines and drugs learned through graph learning, which can be ascribed to three key aspects: (1) the construction of hypergraphs with a hypergraph aggregation mechanism introduces high-order relationships, while the hypergraph fusion method enhances representations of complex multi-relational data across multiple omics; (2) the incorporation of drug-cell association provides supplementary information; and (3) the graph transformer with molecular-specific attention learns more effective representations for drugs.

To demonstrate the potential of integrating multi omics data types for improving drug response prediction, we conduct experiments on different combinations of the three available omics, including single omics (gne, cnv, and mu) and different combinations (gne + cnv, gne + mu, cnv + mu, and gne + cnv + mu). Fig. 4 shows that all multi omics combinations outperform any single omics. Notably, integrating all three omics (gne + cnv + mu) consistently enables HCLGT-DRP to achieve the best performance across two datasets, confirming that multi omics integration enhances prediction performance.

To evaluate the training stability and convergence behavior, we conduct a convergence analysis of HCLGT-DRP. Fig. 5 illustrates the training loss and AUC over epochs on two datasets. The training loss curve exhibits a rapid decline, while the AUC curve shows a swift increase, both converging within a limited number of epochs. Moreover, the absence of significant oscillations or divergence in these curves suggests that the training process is stable.

Table 3
Comparison of model complexity between HCLGT-DRP and RedCDR on two datasets.

Dataset	Model	Average training time per epoch (ms)	Average inference time (ms)	Memory usage (GB)
GDSC	RedCDR	364.57	2.70	2.01
	HCLGT-DRP	558.17	3.83	2.89
CCLE	RedCDR	197.32	2.01	1.70
	HCLGT-DRP	278.15	2.79	1.97

In addition, Table 3 presents a comparison of model complexity between our model and the second-best performing model, RedCDR. The analysis shows that the increased complexity for HCLGT-DRP is primarily due to its more complex model architecture, which incorporates attention interaction mechanisms and graph transformer. These components enhance the model's representation capacity and generalization performance. Considering the improvements in predictive performance, the computational overhead of our model remains within an acceptable and practical range.

5.2. *Predicting response for a new drug or a new cell line (Test 2)*

To showcase the efficacy of HCLGT-DRP in predicting the responses of an unseen cell line or drug, a row or a column in the response matrix is randomly removed as the testing set, while the remaining serve as the training set. Each row in the response matrix denotes a cell line, and each column denotes a drug. The analysis is restricted to cell lines and

Table 4
Performance of different methods in predicting response for a new cell line or a new drug on two datasets. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset		Model	AUC	ACC	Precision	F1 Score
GDSC	New cell line	SRMF	0.5893 ± 4 × 10^{-2}	0.5814 ± 1 × 10^{-3}	0.5637 ± 6 × 10^{-3}	0.6893 ± 1 × 10^{-3}
		MOLI	0.7751 ± 2 × 10^{-3}	0.7246 ± 2 × 10^{-3}	0.7013 ± 4 × 10^{-3}	0.7703 ± 5 × 10^{-3}
		DeepForest	0.7842 ± 3 × 10^{-3}	0.7408 ± 3 × 10^{-3}	0.7219 ± 6 × 10^{-3}	0.7967 ± 8 × 10^{-3}
		DeepCDR	0.7867 ± 6 × 10^{-3}	0.7532 ± 4 × 10^{-3}	0.7382 ± 8 × 10^{-3}	0.7935 ± 2 × 10^{-3}
		GraOmicDRP	0.7319 ± 6 × 10^{-3}	0.7015 ± 2 × 10^{-3}	0.6813 ± 4 × 10^{-3}	0.7519 ± 5 × 10^{-3}
		MOFGCN	0.7209 ± 6 × 10^{-3}	0.6927 ± 1 × 10^{-3}	0.6590 ± 1 × 10^{-3}	0.7572 ± 2 × 10^{-3}
		GraphDRP	0.7019 ± 5 × 10^{-3}	0.6623 ± 2 × 10^{-3}	0.6420 ± 4 × 10^{-3}	0.6915 ± 6 × 10^{-3}
		GraphCDR	0.7210 ± 7 × 10^{-3}	0.6819 ± 5 × 10^{-3}	0.6517 ± 6 × 10^{-3}	0.7430 ± 4 × 10^{-3}
		DGCL	0.7893 ± 4 × 10^{-3}	0.7543 ± 5 × 10^{-3}	0.7339 ± 1 × 10^{-3}	0.7998 ± 4 × 10^{-3}
		RedCDR	0.7963 ± 2 × 10^{-3}	0.7647 ± 6 × 10^{-3}	0.7516 ± 5 × 10^{-3}	0.7906 ± 2 × 10^{-3}
		HCLGT-DRP	**0.8017 ± 4 × 10^{-3}**	**0.7883 ± 1 × 10^{-3}**	**0.7693 ± 3 × 10^{-3}**	**0.8167 ± 5 × 10^{-3}**
	New drug	SRMF	0.6542 ± 2 × 10^{-3}	0.6019 ± 3 × 10^{-3}	0.5864 ± 5 × 10^{-3}	0.6942 ± 1 × 10^{-3}
		MOLI	0.7216 ± 5 × 10^{-3}	0.6619 ± 3 × 10^{-3}	0.6315 ± 3 × 10^{-3}	0.7140 ± 6 × 10^{-3}
		DeepForest	0.7354 ± 6 × 10^{-3}	0.6840 ± 5 × 10^{-3}	0.6519 ± 5 × 10^{-3}	0.7016 ± 2 × 10^{-3}
		DeepCDR	0.7543 ± 1 × 10^{-3}	0.7028 ± 6 × 10^{-3}	0.6895 ± 6 × 10^{-3}	0.7391 ± 3 × 10^{-3}
		GraOmicDRP	0.6746 ± 2 × 10^{-3}	0.6513 ± 6 × 10^{-3}	0.6017 ± 1 × 10^{-2}	0.6578 ± 6 × 10^{-3}
		MOFGCN	0.7501 ± 3 × 10^{-3}	0.6877 ± 7 × 10^{-3}	0.6637 ± 2 × 10^{-3}	0.7413 ± 5 × 10^{-3}
		GraphDRP	0.6423 ± 3 × 10^{-3}	0.6082 ± 2 × 10^{-3}	0.5517 ± 2 × 10^{-3}	0.6374 ± 6 × 10^{-3}
		GraphCDR	0.7703 ± 2 × 10^{-3}	0.7315 ± 1 × 10^{-3}	0.7006 ± 2 × 10^{-3}	0.7654 ± 5 × 10^{-3}
		DGCL	0.7609 ± 4 × 10^{-3}	0.7317 ± 3 × 10^{-3}	0.7119 ± 4 × 10^{-3}	0.7519 ± 5 × 10^{-3}
		RedCDR	0.7892 ± 3 × 10^{-3}	0.7416 ± 3 × 10^{-3}	0.7164 ± 2 × 10^{-3}	0.7593 ± 1 × 10^{-3}
		HCLGT-DRP	**0.7996 ± 8 × 10^{-3}**	**0.7503 ± 5 × 10^{-3}**	**0.7313 ± 6 × 10^{-3}**	**0.7840 ± 2 × 10^{-3}**
CCLE	New cell line	SRMF	0.6073 ± 1 × 10^{-3}	0.7094 ± 3 × 10^{-3}	0.6473 ± 2 × 10^{-3}	0.7750 ± 6 × 10^{-3}
		MOLI	0.8540 ± 2 × 10^{-3}	0.8372 ± 4 × 10^{-3}	0.7348 ± 2 × 10^{-3}	0.8327 ± 5 × 10^{-3}
		DeepForest	0.8607 ± 2 × 10^{-3}	0.8340 ± 6 × 10^{-3}	**0.8074 ± 2 × 10^{-2}**	0.8706 ± 2 × 10^{-3}
		DeepCDR	0.8790 ± 6 × 10^{-3}	0.8605 ± 3 × 10^{-3}	0.8040 ± 2 × 10^{-3}	0.8602 ± 2 × 10^{-3}
		GraOmicDRP	0.7794 ± 8 × 10^{-3}	0.7565 ± 5 × 10^{-3}	0.7169 ± 6 × 10^{-3}	0.7864 ± 5 × 10^{-3}
		MOFGCN	0.8190 ± 3 × 10^{-3}	0.8017 ± 6 × 10^{-3}	0.7673 ± 3 × 10^{-3}	0.8376 ± 2 × 10^{-3}
		GraphDRP	0.7204 ± 3 × 10^{-3}	0.6813 ± 5 × 10^{-2}	0.6073 ± 4 × 10^{-2}	0.7004 ± 3 × 10^{-3}
		GraphCDR	0.7660 ± 2 × 10^{-2}	0.7403 ± 2 × 10^{-3}	0.7052 ± 1 × 10^{-2}	0.7900 ± 2 × 10^{-2}
		DGCL	0.8432 ± 3 × 10^{-3}	0.8149 ± 2 × 10^{-3}	0.7512 ± 3 × 10^{-3}	0.8306 ± 6 × 10^{-3}
		RedCDR	0.8519 ± 2 × 10^{-3}	0.8319 ± 3 × 10^{-3}	0.7583 ± 2 × 10^{-3}	0.8371 ± 4 × 10^{-2}
		HCLGT-DRP	**0.8850 ± 3 × 10^{-3}**	**0.8709 ± 5 × 10^{-3}**	0.8067 ± 1 × 10^{-3}	**0.8796 ± 5 × 10^{-3}**
	New drug	SRMF	0.4861 ± 6 × 10^{-3}	0.5178 ± 3 × 10^{-3}	0.5107 ± 2 × 10^{-3}	0.6759 ± 6 × 10^{-3}
		MOLI	0.7086 ± 4 × 10^{-3}	0.6419 ± 4 × 10^{-3}	0.5910 ± 6 × 10^{-3}	0.7016 ± 3 × 10^{-3}
		DeepForest	0.6375 ± 3 × 10^{-3}	0.6017 ± 3 × 10^{-3}	0.5702 ± 6 × 10^{-3}	0.6233 ± 7 × 10^{-3}
		DeepCDR	0.7317 ± 2 × 10^{-3}	0.6658 ± 5 × 10^{-3}	0.6512 ± 4 × 10^{-3}	0.7298 ± 4 × 10^{-3}
		GraOmicDRP	0.6517 ± 6 × 10^{-3}	0.6017 ± 3 × 10^{-3}	0.5413 ± 2 × 10^{-3}	0.6340 ± 4 × 10^{-3}
		MOFGCN	0.7119 ± 1 × 10^{-3}	0.6577 ± 2 × 10^{-3}	0.6124 ± 2 × 10^{-3}	0.7274 ± 5 × 10^{-3}
		GraphDRP	0.6762 ± 1 × 10^{-3}	0.6513 ± 2 × 10^{-4}	0.6372 ± 2 × 10^{-3}	0.6905 ± 5 × 10^{-3}
		GraphCDR	**0.7514 ± 2 × 10^{-3}**	0.6801 ± 4 × 10^{-2}	0.6517 ± 3 × 10^{-2}	0.7310 ± 5 × 10^{-3}
		DGCL	0.7319 ± 3 × 10^{-3}	0.6518 ± 4 × 10^{-2}	0.6007 ± 2 × 10^{-2}	0.7011 ± 2 × 10^{-3}
		RedCDR	0.7008 ± 4 × 10^{-3}	0.6919 ± 7 × 10^{-2}	0.6513 ± 4 × 10^{-2}	0.6987 ± 5 × 10^{-3}
		HCLGT-DRP	0.7363 ± 2 × 10^{-3}	**0.7029 ± 6 × 10^{-3}**	**0.6712 ± 3 × 10^{-2}**	**0.7312 ± 5 × 10^{-3}**

drugs with at least 10 sensitive samples to avoid selecting overly general or particular drugs (Staunton et al., 2001). After processing, the GDSC dataset comprises 658 cell lines and 227 drugs, and the CCLE dataset includes 26 cell lines and 20 drugs. Five-fold cross-validation is conducted and repeated five times.

The results illustrated in Table 4 show that HCLGT-DRP outperforms baselines in most cases across all evaluation metrics on both datasets, except for ranking second in terms of AUC value when predicting

Table 5
Performance of different methods for predicting a single drug response on two datasets. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Model	AUC	ACC	Precision	F1 Score
GDSC	SRMF	$0.6554 \pm 2 \times 10^{-3}$	$0.5890 \pm 3 \times 10^{-3}$	$0.5748 \pm 1 \times 10^{-3}$	$0.6308 \pm 2 \times 10^{-3}$
	MOLI	$0.6842 \pm 5 \times 10^{-3}$	$0.6833 \pm 5 \times 10^{-3}$	$0.6521 \pm 3 \times 10^{-3}$	$0.7410 \pm 4 \times 10^{-2}$
	DeepForest	$0.6149 \pm 2 \times 10^{-3}$	$0.6517 \pm 5 \times 10^{-3}$	$0.6190 \pm 5 \times 10^{-3}$	$0.6993 \pm 4 \times 10^{-3}$
	DeepCDR	$0.7340 \pm 4 \times 10^{-2}$	$0.7258 \pm 3 \times 10^{-3}$	$0.6829 \pm 2 \times 10^{-3}$	$0.7580 \pm 3 \times 10^{-3}$
	GraOmicDRP	$0.7043 \pm 2 \times 10^{-2}$	$0.7269 \pm 4 \times 10^{-2}$	$0.6879 \pm 2 \times 10^{-2}$	$0.7790 \pm 2 \times 10^{-3}$
	MOFGCN	$0.8331 \pm 2 \times 10^{-2}$	$\underline{0.8042 \pm 2 \times 10^{-3}}$	$0.7793 \pm 3 \times 10^{-2}$	$\underline{0.8351 \pm 2 \times 10^{-3}}$
	GraphDRP	$0.7043 \pm 2 \times 10^{-2}$	$0.7269 \pm 4 \times 10^{-2}$	$0.6879 \pm 2 \times 10^{-2}$	$0.7790 \pm 2 \times 10^{-3}$
	GraphCDR	$0.7235 \pm 1 \times 10^{-2}$	$0.6987 \pm 5 \times 10^{-2}$	$0.6842 \pm 4 \times 10^{-3}$	$0.7208 \pm 5 \times 10^{-3}$
	DGCL	$0.7458 \pm 3 \times 10^{-3}$	$0.7032 \pm 2 \times 10^{-3}$	$0.7143 \pm 3 \times 10^{-3}$	$0.7422 \pm 5 \times 10^{-3}$
	RedCDR	$\underline{0.8343 \pm 6 \times 10^{-3}}$	$0.8007 \pm 3 \times 10^{-3}$	$\underline{0.7824 \pm 8 \times 10^{-3}}$	$0.8349 \pm 5 \times 10^{-3}$
	HCLGT-DRP	$\mathbf{0.8677 \pm 5 \times 10^{-3}}$	$\mathbf{0.8263 \pm 5 \times 10^{-3}}$	$\mathbf{0.7942 \pm 2 \times 10^{-3}}$	$\mathbf{0.8483 \pm 7 \times 10^{-3}}$
CCLE	SRMF	$0.7404 \pm 1 \times 10^{-3}$	$0.6697 \pm 3 \times 10^{-3}$	$0.5933 \pm 5 \times 10^{-3}$	$0.6543 \pm 4 \times 10^{-3}$
	MOLI	$0.6690 \pm 2 \times 10^{-3}$	$0.6689 \pm 2 \times 10^{-3}$	$0.6570 \pm 1 \times 10^{-3}$	$0.7247 \pm 3 \times 10^{-3}$
	DeepForest	$0.6044 \pm 3 \times 10^{-3}$	$0.6017 \pm 1 \times 10^{-3}$	$0.6190 \pm 5 \times 10^{-3}$	$0.7244 \pm 3 \times 10^{-3}$
	DeepCDR	$0.6953 \pm 1 \times 10^{-2}$	$0.6710 \pm 1 \times 10^{-3}$	$0.6533 \pm 4 \times 10^{-3}$	$0.7459 \pm 3 \times 10^{-3}$
	GraOmicDRP	$0.6623 \pm 4 \times 10^{-2}$	$0.6858 \pm 3 \times 10^{-2}$	$0.6544 \pm 1 \times 10^{-3}$	$0.7074 \pm 3 \times 10^{-2}$
	MOFGCN	$0.8423 \pm 3 \times 10^{-2}$	$0.8042 \pm 4 \times 10^{-2}$	$\underline{0.7813 \pm 3 \times 10^{-3}}$	$0.8252 \pm 2 \times 10^{-3}$
	GraphDRP	$0.7267 \pm 2 \times 10^{-2}$	$0.7013 \pm 2 \times 10^{-2}$	$0.6528 \pm 3 \times 10^{-2}$	$0.7763 \pm 3 \times 10^{-2}$
	GraphCDR	$0.7756 \pm 2 \times 10^{-2}$	$0.7089 \pm 2 \times 10^{-3}$	$0.6957 \pm 1 \times 10^{-2}$	$0.7860 \pm 2 \times 10^{-2}$
	DGCL	$0.7541 \pm 2 \times 10^{-3}$	$0.7059 \pm 3 \times 10^{-3}$	$0.6547 \pm 5 \times 10^{-3}$	$0.7133 \pm 2 \times 10^{-3}$
	RedCDR	$\underline{0.8573 \pm 5 \times 10^{-3}}$	$\underline{0.8074 \pm 7 \times 10^{-3}}$	$0.7533 \pm 2 \times 10^{-3}$	$\underline{0.8274 \pm 5 \times 10^{-3}}$
	HCLGT-DRP	$\mathbf{0.8603 \pm 3 \times 10^{-3}}$	$\mathbf{0.8147 \pm 5 \times 10^{-3}}$	$\mathbf{0.7833 \pm 3 \times 10^{-3}}$	$\mathbf{0.8560 \pm 6 \times 10^{-3}}$

responses for new drugs and in precision value when predicting responses for new cell lines, both on the CCLE dataset. However, HCLGT-DRP consistently performs best on the overall performance, demonstrating that our model remains competitive.

5.3. Predicting a single drug response (Test 3)

To evaluate the ability of HCLGT-DRP to predict the response of a single drug across multiple cell lines, a five-fold cross-validation is performed and repeated five times for each drug. Similar to Test 2, the focus is on drugs that have over 10 sensitive samples.

As depicted in Table 5, HCLGT-DRP still controls the best performance among all methods, demonstrating a remarkable ability to predict single drug sensitivity. Specifically, on the GDSC dataset, HCLGT-DRP achieves an AUC of 0.8677, ACC of 0.8263, precision of 0.7942, and F1 of 0.8483, surpassing the second-best method by 3.34 %, 2.21 %, 1.18 %, and 1.32 %, respectively. On the CCLE dataset, HCLGT-DRP shows improvements of 3.00 %, 0.73 %, 0.20 %, and 2.86 % in AUC, ACC, precision and F1, respectively, achieving values of 0.8603, 0.8147, 0.7833, and 0.8560, compared to the second-best method.

Table 6
Results of ablation studies on two datasets. The best results are highlighted in bold.

Dataset	Method	AUC	ACC	Precision	F1 Score
GDSC	w/o HRL	$*0.8521 \pm 5 \times 10^{-5}$	$*0.7857 \pm 5 \times 10^{-5}$	$*0.7841 \pm 2 \times 10^{-5}$	$*0.8240 \pm 6 \times 10^{-5}$
	w/o HGA	$*0.8642 \pm 3 \times 10^{-5}$	$*0.8006 \pm 2 \times 10^{-5}$	$*0.7641 \pm 4 \times 10^{-5}$	$*0.8216 \pm 4 \times 10^{-5}$
	w/o HF	$*0.8613 \pm 2 \times 10^{-5}$	$*0.7941 \pm 3 \times 10^{-5}$	$*0.7917 \pm 4 \times 10^{-5}$	$*0.8208 \pm 3 \times 10^{-5}$
	w/o ARL	$*0.8602 \pm 6 \times 10^{-5}$	$*0.7937 \pm 5 \times 10^{-5}$	$*0.7906 \pm 3 \times 10^{-5}$	$*0.8298 \pm 4 \times 10^{-5}$
	w/o MSA	$*0.8740 \pm 5 \times 10^{-5}$	$*0.8023 \pm 3 \times 10^{-5}$	$0.7942 \pm 2 \times 10^{-5}$	$*0.8215 \pm 6 \times 10^{-5}$
	−concat	$*0.8732 \pm 2 \times 10^{-5}$	$*0.8019 \pm 5 \times 10^{-5}$	$*0.7913 \pm 3 \times 10^{-5}$	$*0.8297 \pm 1 \times 10^{-5}$
	−add	$*0.8740 \pm 2 \times 10^{-4}$	$0.8021 \pm 1 \times 10^{-4}$	$*0.7919 \pm 2 \times 10^{-4}$	$0.8302 \pm 3 \times 10^{-5}$
	−atten	$*0.8743 \pm 3 \times 10^{-4}$	$*0.8020 \pm 3 \times 10^{-4}$	$0.7930 \pm 6 \times 10^{-4}$	$*0.8295 \pm 2 \times 10^{-5}$
	HCLGT-DRP	$\mathbf{0.8796 \pm 4 \times 10^{-5}}$	$\mathbf{0.8034 \pm 2 \times 10^{-5}}$	$\mathbf{0.7961 \pm 3 \times 10^{-5}}$	$\mathbf{0.8317 \pm 4 \times 10^{-5}}$
CCLE	w/o HRL	$*0.8642 \pm 3 \times 10^{-5}$	$*0.8012 \pm 2 \times 10^{-5}$	$*0.7819 \pm 3 \times 10^{-5}$	$*0.8059 \pm 2 \times 10^{-5}$
	w/o HGA	$*0.8651 \pm 1 \times 10^{-5}$	$*0.8031 \pm 2 \times 10^{-5}$	$*0.7805 \pm 4 \times 10^{-5}$	$*0.8062 \pm 2 \times 10^{-5}$
	w/o HF	$*0.8653 \pm 3 \times 10^{-5}$	$*0.8005 \pm 2 \times 10^{-5}$	$*0.7813 \pm 2 \times 10^{-5}$	$*0.8091 \pm 3 \times 10^{-5}$
	w/o ARL	$*0.8617 \pm 4 \times 10^{-5}$	$*0.7946 \pm 4 \times 10^{-5}$	$*0.7794 \pm 5 \times 10^{-5}$	$*0.8067 \pm 1 \times 10^{-5}$
	w/o MSA	$*0.8662 \pm 2 \times 10^{-5}$	$*0.8022 \pm 5 \times 10^{-5}$	$*0.7814 \pm 5 \times 10^{-5}$	$0.8019 \pm 2 \times 10^{-5}$
	−concat	$0.8641 \pm 5 \times 10^{-5}$	$*0.7961 \pm 3 \times 10^{-5}$	$*0.7822 \pm 4 \times 10^{-5}$	$*0.8081 \pm 3 \times 10^{-5}$
	−add	$0.8684 \pm 4 \times 10^{-5}$	$*0.8029 \pm 1 \times 10^{-4}$	$0.7851 \pm 2 \times 10^{-4}$	$*0.8033 \pm 3 \times 10^{-4}$
	−atten	$*0.8672 \pm 1 \times 10^{-4}$	$\mathbf{0.8052 \pm 4 \times 10^{-5}}$	$*0.7827 \pm 1 \times 10^{-5}$	$0.8096 \pm 1 \times 10^{-5}$
	HCLGT-DRP	$\mathbf{0.8703 \pm 1 \times 10^{-5}}$	$0.8046 \pm 3 \times 10^{-5}$	$\mathbf{0.7854 \pm 3 \times 10^{-5}}$	$\mathbf{0.8103 \pm 2 \times 10^{-5}}$

* denotes p-value < 0.05.

5.4. Ablation studies (Test 4)

To assess the contribution of different components to the model's performance, ablation studies are conducted using randomly zeroing cross-validation on two datasets by comparing the following six variants of HCLGT-DRP:

- **w/o HRL**: removes the hypergraph representation learning for cell lines and instead employs GNN for learning on a basic graph structure of cell lines.
- **w/o HGA**: removes the hypergraph aggregation mechanism used in hypergraph construction.
- **w/o HF**: removes the hypergraph fusion mechanism and utilizes an averaging operation to fuse representations.
- **w/o ARL**: removes the cell line-drug association graph representation learning in cell line feature extraction.
- **w/o MSA**: removes the molecular-specific attention layer within the graph transformer.

To further validate the effectiveness of the proposed contrastive learning strategy, we replaced it with other basic fusion methods:

- **-concat**: directly concatenates the representation from the hypergraph learning and association graph learning.
- **-add**: directly sums representations from hypergraph learning and association graph learning.
- **-atten:** uses a learnable mapping layer to obtain the attention scores of the representations from hypergraph learning and association graph learning, followed by a weighted fusion.

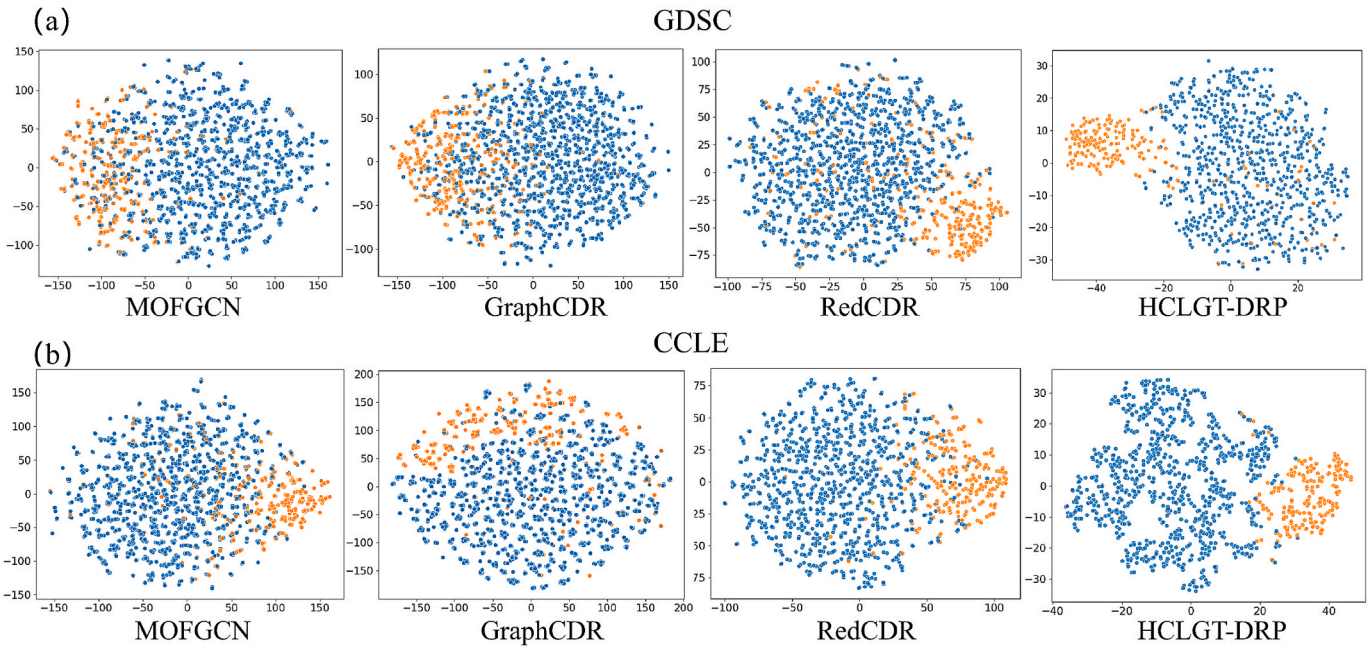


Fig. 6. T-SNE visualization of representations for four methods on the two datasets. (a) Visualization results of GDSC dataset. (b) Visualization results of CCLE dataset.

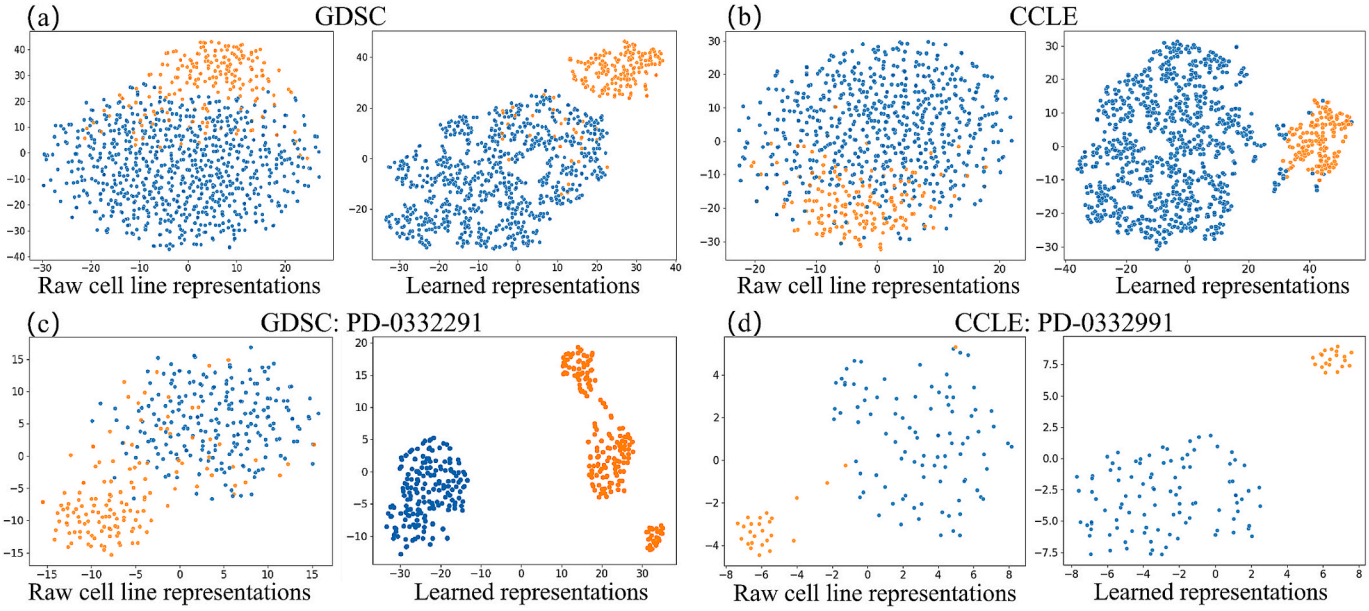


Fig. 7. T-SNE visualization of the classification space learned by HCLGT-DRP on the GDSC and CCLE datasets. (a)–(b) Visualization results of raw cell line representations and learned representations. (c)–(d) Visualization results of the distribution of cell line representations affected by the targeted drug PD-0332991.

Table 6 shows that HCLGT-DRP significantly (p-value < 0.05, paired t-test) surpasses the variants across the most metrics, demonstrating that each component positively enhances the final performance. The results of the variant (w/o HRL) emphasize the importance of introducing the hypergraph structure for omics-specific representation learning. The variant (w/o HGA) indicates that the hypergraph aggregation mechanism optimizes the overall structure of the hypergraph while retaining key information, enhancing predictive performance. The variant (w/o HF) indicates that the hypergraph fusion mechanism effectively integrates node and hyperedge information from various omics hypergraphs, capturing complex multi-relational data structures. The variant's performance (w/o ARL) highlights the critical contribution of the cell line-drug association graph in extracting cell line features. The results of the variant (w/o MSA) emphasize the critical role of molecular-specific attention in extracting drug features, which considers not only the drug attributes but also their structural information, enhancing drug representation. By comparing several fusion methods, the proposed contrastive learning approach outperforms other fusion methods in most cases across all evaluation metrics on both datasets, demonstrating its superior ability in feature fusion.

5.5. *Visualization of feature representation (Test 5)*

To intuitively evaluate the overall classification performance of

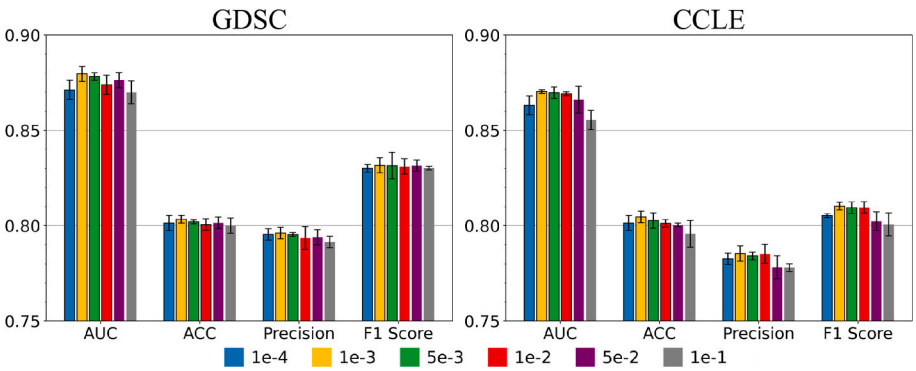


Fig. 8. Performance comparisons under different values of *lr* on the GDSC and CCLE datasets.

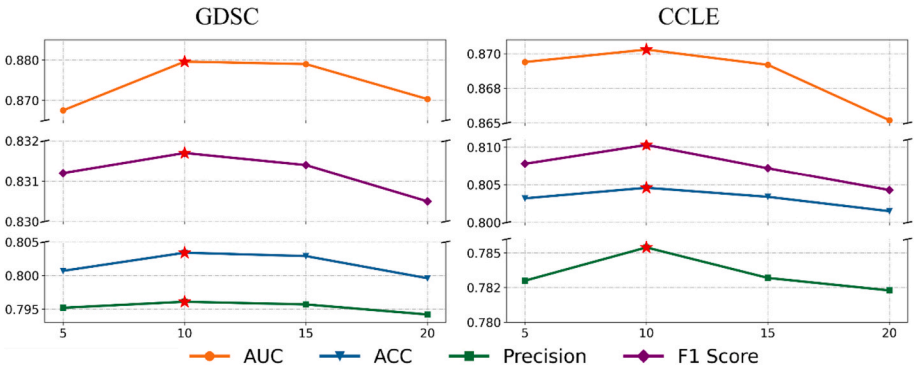


Fig. 9. Performance comparisons under different values of *K* on the GDSC and CCLE datasets. Red stars indicate the optimal hyper-parameter values determined by the AUC, ACC, Precision and F1.

HCLGT-DRP, we employ t-SNE (Maaten & Hinton, 2008) to perform dimensionality reduction visualization analysis for comparing the learned representations used for the final classification task. We select three representative baselines, MOFGCN, GraphCDR, and RedCDR, for comparison with the HCLGT-DRP. The data is generated by concatenating the learned cell lines representations with the learned drug representations. As shown in Fig. 6, HCLGT-DRP achieves better intra-class compactness and inter-class separability across two datasets, indicating that the hypergraph contrastive learning and graph transformer architectures can be effectively utilized for DRP.

To further investigate the capability of HCLGT-DRP to extract and integrate features from multi omics data, we compare the raw cell line data with the representations learned by HCLGT-DRP, and visualize the results using t-SNE. Fig. 7 (a)–(b) illustrate the distributions of the raw cell line representations and the learned representations. In addition, to analyze the ability of HCLGT-DRP to identify cell line features associated with susceptibility to targeted therapies, PD-0332991 (Peng, Liu, et al., 2022), a CDK4-targeted drug (Staunton et al., 2001), is selected from two datasets and the changes in cell line embeddings before and after training are visualized in Fig. 7 (c)–(d).

Fig. 7 shows that the raw cell line data are tightly grouped, presenting difficulties for accurately predicting drug responses. In contrast, the fused representations obtained after applying HCLGT-DRP demonstrate enhanced discriminatory ability. This observation suggests that the representations learned by HCLGT-DRP may implicitly encode resistance or sensitivity labels, which could be leveraged for drug response prediction task.

5.6. *Influence of key hyper-parameters (Test 6)*

This section conducts experiments to investigate the impacts of four key hyper-parameters on the performance of HCLGT-DRP, using

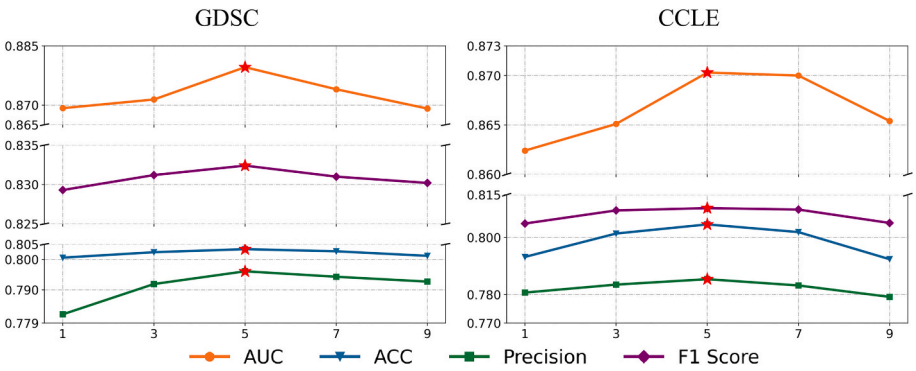


Fig. 10. Performance comparisons under different values of *k* on the GDSC and CCLE datasets. Red stars indicate the optimal hyper-parameter values determined by the AUC, ACC, Precision and F1.

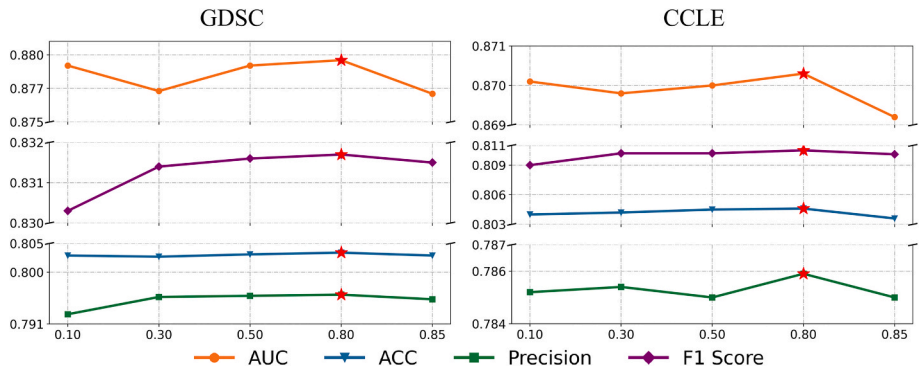


Fig. 11. Performance comparisons under different values of on the GDSC and CCLE datasets. Red stars indicate the optimal hyper-parameter values according to AUC, ACC, Precision and F1.

Table 7
Performance of different methods for predicting drug response on TCGA datasets. The best and second-best results are highlighted in bold and underlined, respectively.

Dataset	Model	AUC	ACC	Precision	F1 Score
TCGA	SRMF	$0.4146 \pm 4 \times 10^{-5}$	$0.3610 \pm 3 \times 10^{-5}$	$0.3571 \pm 5 \times 10^{-5}$	$0.4313 \pm 2 \times 10^{-4}$
	MOLI	$0.5018 \pm 2 \times 10^{-3}$	$0.4791 \pm 4 \times 10^{-3}$	$0.4503 \pm 5 \times 10^{-3}$	$0.5080 \pm 4 \times 10^{-3}$
	DeepForest	$0.4602 \pm 4 \times 10^{-3}$	$0.4041 \pm 3 \times 10^{-4}$	$0.3874 \pm 3 \times 10^{-3}$	$0.4802 \pm 2 \times 10^{-3}$
	DeepCDR	$0.6421 \pm 2 \times 10^{-3}$	$0.5863 \pm 5 \times 10^{-3}$	$0.5730 \pm 4 \times 10^{-3}$	$0.6617 \pm 3 \times 10^{-3}$
	GraOmicDRP	$0.6347 \pm 3 \times 10^{-3}$	$0.5321 \pm 3 \times 10^{-3}$	$0.5514 \pm 5 \times 10^{-3}$	$0.6015 \pm 4 \times 10^{-3}$
	MOFGCN	$0.5391 \pm 4 \times 10^{-3}$	$0.4211 \pm 2 \times 10^{-3}$	$0.4071 \pm 3 \times 10^{-3}$	$0.5514 \pm 5 \times 10^{-3}$
	GraphDRP	$0.5609 \pm 3 \times 10^{-3}$	$0.5316 \pm 5 \times 10^{-4}$	$0.5195 \pm 1 \times 10^{-3}$	$0.6204 \pm 2 \times 10^{-4}$
	GraphCDR	$0.6635 \pm 1 \times 10^{-3}$	**$0.6431 \pm 5 \times 10^{-2}$**	$0.6139 \pm 4 \times 10^{-3}$	$0.7079 \pm 5 \times 10^{-3}$
	DGCL	$0.6517 \pm 4 \times 10^{-3}$	$0.6014 \pm 1 \times 10^{-3}$	$0.5634 \pm 3 \times 10^{-3}$	$0.6750 \pm 2 \times 10^{-3}$
	RedCDR	$0.6638 \pm 5 \times 10^{-3}$	$0.6236 \pm 2 \times 10^{-3}$	$0.5871 \pm 2 \times 10^{-3}$	$0.6940 \pm 4 \times 10^{-3}$
	HCLGT-DRP	**$0.7029 \pm 3 \times 10^{-3}$**	$0.6342 \pm 1 \times 10^{-3}$	**$0.6266 \pm 2 \times 10^{-3}$**	**$0.7283 \pm 7 \times 10^{-3}$**

randomly zeroing cross-validation on two datasets.

The first hyper-parameter is learning rate (*lr*). The optimal value of *lr* is searched from the set {1e-4, 1e-3, 5e-3, 1e-2, 5e-2, 1e-1}. Fig. 8 indicates that the optimal *lr* for both datasets is 1e-3. When *lr* is excessively high, it causes unstable training and prevents the model from converging properly, resulting in suboptimal performance. A moderate learning rate strikes a balance between stable optimization and fast convergence.

The second hyper-parameter is *K*, determines the number of top-K nearest neighbors used to construct each hyperedge in the hypergraph. It plays a critical role in shaping the structure of the hypergraph. We empirically evaluate the optimal value from the set {5, 10, 15, 20}. As shown in Fig. 9, the best performance on both datasets is achieved when K = 10. A smaller *K* leads to a sparser hypergraph that overly emphasizes local neighborhood information, resulting in insufficient information propagation. In contrast, a larger *K* may introduce additional noise from less relevant neighbors, making the model prone to overfitting.

The third hyper-parameter is *k*, indicating the walk step size under hypergraph aggregation in Equation (12), which affects the complex interactions between nodes and hyperedges across the hypergraph. Fig. 10 shows that the model performs optimally when the step size *k* is set to 5. As *k* decreases, the model shifts its attention toward individual similar cell lines, where each step of information propagation is limited to nearby neighbors. Conversely, when *k* is too large, irrelevant noise is introduced, and over-smoothing of the information can occur, which degrades the model's performance.

The fourth hyper-parameter is λ , which refers to the trade-off parameter between L_{HGF} and L_{AGF} in Equation (25). Adjusting λ facilitates a balance of the importance between the hypergraph representation learning and the association graph representation learning, ensuring that both learned representations contribute appropriately to the overall model training. Fig. 11 shows that the model performs best at $\lambda = 0.8$ across both datasets. The model performance can degrade when λ is too small or too large. Both datasets exhibit a comparable sensitivity to variations in this parameter.

5.7. External validation on in vivo dataset (Test 7)

External validation serves as a critical step to evaluate the generalization ability and real-world applicability of a DRP model. To evaluate the transferability of the model from *in vitro* to *in vivo* dataset, HCLGT-DRP and several baseline models were trained on *in vitro* dataset GDSC in randomly zeroing cross-validation and subsequently applied to predict drug responses on *in vivo* dataset TCGA.

Table 8
Top 10 predicted sensitive cell lines for Dasatinib and GSK690693.

Drug	Rank	Cell line	PMID	Drug	Rank	Cell line	PMID
Dasatinib	1	SW982	N/A	GSK690693	1	GA-10	N/A
	2	SW-1710	N/A		2	**RCH-ACV**	19064730
	3	**MCF-7**	22306341		3	KYAE-1	N/A
	4	OS-RC-2	N/A		4	GESS	N/A
	5	SNG-M	N/A		5	**REH**	19064730
	6	**HEY**	20629079		6	IGROV1	N/A
	7	NCI-H196	N/A		7	HC-1	N/A
	8	**Cal12T**	22649091		8	**Molt-16**	19064730
	9	NCI-H1105	N/A		9	MM1-S	N/A
	10	ACHN	N/A		10	RH-18	N/A

Table 7 shows that HCLGT-DRP consistently achieved the best overall performance, obtaining the highest AUC of 0.7029. Although its ACC was slightly lower than the second-best baseline, HCLGT-DRP demonstrates superior performance across most evaluation metrics. These results confirm the generalization ability of HCLGT-DRP from *in vitro* cell line data to *in vivo* clinical data.

5.8. Case study of predicting unknown responses (Test 8)

Similar to previous studies (X. Liu et al., 2022; Peng, Chen, et al., 2022; Peng, Liu, et al., 2022; Song et al., 2025), we conduct a case study to further validate the predictive capability of HCLGT-DRP for novel cell line-drug responses. HCLGT-DRP is trained using all known drug responses from the GDSC dataset and applied to predict unknown responses. For the analysis, the focus is placed on two clinically approved drugs: Dasatinib and GSK690693.

Table 8 lists the top 10 cell lines predicted by HCLGT-DRP to exhibit sensitivity to Dasatinib and GSK690693. Since the true values for these unknown responses are unavailable, we validate the predictions by referencing relevant literature. For both Dasatinib and GSK690693, three cell lines are confirmed to exhibit sensitivity. Park et al. (Park et al., 2012) reported that the MCF-7 cell line, derived from invasive breast ductal carcinoma, is moderately sensitive to Dasatinib. Le et al. (Le et al., n.d.) demonstrated that Dasatinib inhibited tumor growth and induced both autophagy and apoptosis in HEY ovarian cancer cells. Sen et al. (Sen et al., 2012) identified two NSCLC cell lines, H1666 and Cal12T, as sensitive to Dasatinib. Levy et al. (Levy et al., 2009) explored the effectiveness of GSK690693 in inhibiting cell proliferation in T-cell and B-/pre-B-cell origin cells within the ALL cell panel, finding that RCH-ACV, REH, and Molt-16 exhibited sensitivity to GSK690693. This case study demonstrates the capability of HCLGT-DRP to identify novel drug-cell line responses, emphasizing its potential for predicting unknown drug sensitivities.

5.9. Limitations and future work

Despite the promising results achieved by HCLGT-DRP, several limitations persist. First, only three types of omics data are considered in our work. The potential impact of integrating additional omics profiles of cell lines or data from other modalities on performance remains uncertain. Future researches can investigate the feasibility of incorporating other omics data or modalities to evaluate the generalization capabilities of HCLGT-DRP (Y. Li et al., 2023). Second, this study utilizes only cell line-drug association to provide auxiliary information for cell line representation learning. Exploring diverse biological association information, such as drug-gene interactions (Tao et al., 2023) and target protein identification (Ha et al., 2021), as supplementary support to enhance representations of drugs and cell lines represents a promising direction for future researches. Third, the current approach for drug feature extraction solely relies on the molecular graph. Future efforts can explore combining other drug representations, such as 3D structure information of drugs (Y. Zhang et al., 2024), and consider employing molecular pre-training strategies to achieve more accurate drug representations (Zhu et al., 2021). Lastly, its biological interpretability remains limited. Future studies can focus on enhancing interpretability by incorporating biologically meaningful priors into the model architecture, such as known signaling pathways (Y. Wang et al., 2024) and gene ontologies(Gao et al., 2025).

6. Conclusions

In this study, we propose a novel model named HCLGT-DRP for predicting drug response. By leveraging hypergraph structures to model high-order relationships within single-omics data of cell lines, incorporating cell line-drug association and utilizing an enhanced graph transformer, HCLGT-DRP effectively addresses the limitations of existing approaches. Extensive experiments on two publicly available datasets and external validation on *in vivo* dataset showcase that HCLGT-DRP outperforms ten outstanding baselines across various DRP tasks, highlighting its great potential for drug response prediction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments. This work was supported in part by the National Science and Technology Major Project of China (2024ZD0523903), the National Natural Science Foundation of China (72342014), the Key Research and Development Program of Sichuan Province (2024YFHZ0060), and the Key Research Project of Health Information Association of Sichuan Province (2024028).

Data availability

Datasets used in this study are available at https://github.com/weiba/MOFGCN. The source code of HCLGT-DRP will be released at https://github.com/wenz-uestc/HCLGT-DRP upon publication.

References

Baptista, D., Ferreira, P. G., & Rocha, M. (2021). Deep learning for drug response prediction in cancer. *Briefings in Bioinformatics, 22*(1), 360–379. https://doi.org/10.1093/bib/bbz171

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., & Garraway, L. A. (2012). CCLE:The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature, 483*(7391), 603–607. https://doi.org/10.1038/nature11003

Bento, A. P., Hersey, A., Félix, E., Landrum, G., Gaulton, A., Atkinson, F., Bellis, L. J., De Veij, M., & Leach, A. R. (2020). RDKit: An open source chemical structure curation pipeline using RDKit. *Journal of Cheminformatics, 12*(1), 51. https://doi.org/10.1186/s13321-020-00456-1

Cancer. (n.d.). Retrieved January 20, 2025, from https://www.who.int/news-room/fact-sheets/detail/cancer.

Chu, T., Nguyen, T. T., Hai, B. D., Nguyen, Q. H., & Nguyen, T. (2023). GraTransDRP: Graph Transformer for Drug Response Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20*(2), 1065–1072. https://doi.org/10.1109/TCBB.2022.3206888

Ding, Z., Zu, S., & Gu, J. (2016). Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics, 32*(19), 2891–2895. https://doi.org/10.1093/bioinformatics/btw344

Dong, Z., Zhang, N., Li, C., Wang, H., Fang, Y., Wang, J., & Zheng, X. (2015). Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer, 15*(1), 489. https://doi.org/10.1186/s12885-015-1492-6

Duvenaud D. K., Maclaurin D., Iparraguirre J., Bombarell R., Hirzel T., Aspuru-Guzik A., & Adams R. P. (2015). Convolutional Networks on Graphs for Learning Molecular Fingerprints. Advances in Neural Information Processing Systems, 28. https://proceedings.neurips.cc/paper_files/paper/2015/hash/f9be311e65d81a9ad8150a60844bb94c-Abstract.html.

Gao, Q., Xu, T., Li, X., Gao, W., Shi, H., Zhang, Y., Chen, J., & Yue, Z. (2025). Interpretable Dynamic Directed Graph Convolutional Network for Multi-Relational Prediction of Missense Mutation and Drug Response. *IEEE Journal of Biomedical and Health Informatics, 29*(2), 1514–1524. https://doi.org/10.1109/JBHI.2024.3483316

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., Greninger, P., Thompson, I. R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R. J., Bignell, G. R., Tam, A. T., Davies, H., Stevenson, J. A., & Benes, C. H. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature, 483*(7391), 570–575. https://doi.org/10.1038/nature11005

Ha, J., Park, H., Park, J., & Park, S. B. (2021). Recent advances in identifying protein targets in drug discovery. *Cell Chemical Biology, 28*(3), 394–423. https://doi.org/10.1016/j.chembiol.2020.12.001

Hu, X., Zhang, P., Zhang, J., & Deng, L. (2024). DeepFusionCDR: Employing Multi-Omics Integration and Molecule-Specific Transformers for Enhanced Prediction of Cancer Drug Responses. *IEEE Journal of Biomedical and Health Informatics, 28*(10), 6248–6258. https://doi.org/10.1109/JBHI.2024.3417014

Jiang J., Wei Y., Feng Y., Cao J., & Gao Y. (2019). Dynamic Hypergraph Neural Networks. 2635–2641.

Jin, S., Hong, Y., Zeng, L., Jiang, Y., Lin, Y., Wei, L., Yu, Z., Zeng, X., & Liu, X. (2023a). A general hypergraph learning algorithm for drug multi-task predictions in micro-to-macro biomedical networks. *PLOS Computational Biology, 19*(11), Article e1011597. https://doi.org/10.1371/journal.pcbi.1011597

Jin, S., Hong, Y., Zeng, L., Jiang, Y., Lin, Y., Wei, L., Yu, Z., Zeng, X., & Liu, X. (2023b). HGDrug: A general hypergraph learning algorithm for drug multi-task predictions in micro-to-macro biomedical networks. *PLOS Computational Biology, 19*(11), Article e1011597. https://doi.org/10.1371/journal.pcbi.1011597

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research, 47*(D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033

Le, X.-F., Mao, W., Lu, Z., Carter, B. Z., & Bast, R. C. (n.d.). Dasatinib induces autophagic cell death in human ovarian cancer. doi: 10.1002/cncr.25426.

Levy, D. S., Kahana, J. A., & Kumar, R. (2009). AKT inhibitor, GSK690693, induces growth inhibition and apoptosis in acute lymphoblastic leukemia cell lines. *Blood, 113*(8), 1723–1729. https://doi.org/10.1182/blood-2008-02-137737

Li, M., Wang, Y., Zheng, R., Shi, X., Li, Y., Wu, F.-X., & Wang, J. (2021). DeepDSC: A Deep Learning Method to Predict Drug Sensitivity of Cancer Cell Lines. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18*(2), 575–582. https://doi.org/10.1109/TCBB.2019.2919581

Li, Y., Guo, Z., Gao, X., & Wang, G. (2023). MMCL-CDR: Enhancing cancer drug response prediction with multi-omics and morphology images contrastive representation learning. *Bioinformatics, 39*(12), Article btad734. https://doi.org/10.1093/bioinformatics/btad734

Lin, C.-X., Guan, Y., & Li, H.-D. (2024). Artificial intelligence approaches for molecular representation in drug response prediction. *Current Opinion in Structural Biology, 84*, Article 102747. https://doi.org/10.1016/j.sbi.2023.102747

Liu, J., Guan, S., Zou, Q., Wu, H., Tiwari, P., & Ding, Y. (2024). AMDGT: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction. *Knowledge-Based Systems, 284*, Article 111329. https://doi.org/10.1016/j.knosys.2023.111329

Liu, Q., Hu, Z., Jiang, R., & Zhou, M. (2020). DeepCDR: A hybrid graph convolutional network for predicting cancer drug response. Bioinformatics, 36(Supplement_2), i911–i918. doi: 10.1093/bioinformatics/btaa822.

Liu, X., Song, C., Huang, F., Fu, H., Xiao, W., & Zhang, W. (2022). GraphCDR: A graph neural network method with contrastive learning for cancer drug response prediction. *Briefings in Bioinformatics, 23*(1), Article bbab457. https://doi.org/10.1093/bib/bbab457

Liu, Y., Tong, S., & Chen, Y. (2023). HMM-GDAN: Hybrid multi-view and multi-scale graph duplex-attention networks for drug response prediction in cancer. *Neural Networks, 167*, 213–222. https://doi.org/10.1016/j.neunet.2023.08.036

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*(86), 2579–2605.

Nguyen, G. T. T., Vu, H. D., & Le, D.-H. (2022). GraOmicDRP: Integrating Molecular Graph Data of Drugs and Multiple -Omic Data of Cell Lines for Drug Response Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19*(2), 710–717. https://doi.org/10.1109/TCBB.2021.3096960

Nguyen, T., Nguyen, G. T. T., Nguyen, T., & Le, D.-H. (2022). GraphDRP:Graph Convolutional Networks for Drug Response Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19*(1), 146–154. https://doi.org/10.1109/TCBB.2021.3060430

van den Oord, A., Li, Y., & Vinyals, O. (2019). *Representation Learning with Contrastive Predictive Coding, No. arXiv:1807.03748*, arXiv. https://doi.org/10.48550/arXiv.1807.03748

Park, B. J., Whichard, Z. L., & Corey, S. J. (2012). Dasatinib synergizes with both cytotoxic and signal transduction inhibitors in heterogeneous breast cancer cell lines—Lessons for design of combination targeted therapy. *Cancer Letters, 320*(1), 104–110. https://doi.org/10.1016/j.canlet.2012.01.039

Partin, A., Brettin, T. S., Zhu, Y., Narykov, O., Clyde, A., Overbeek, J., & Stevens, R. L. (2023). Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Frontiers in Medicine, 10*, Article 1086097. https://doi.org/10.3389/fmed.2023.1086097

Peng, W., Chen, T., & Dai, W. (2022). MOFGCN: Predicting Drug Response Based on Multi-Omics Fusion and Graph Convolution. *IEEE Journal of Biomedical and Health Informatics, 26*(3), 1384–1393. https://doi.org/10.1109/JBHI.2021.3102186

Peng, W., Chen, T., Liu, H., Dai, W., Yu, N., & Lan, W. (2023). TSGCNN :Improving drug response prediction based on two-space graph convolution. *Computers in Biology and Medicine, 158*, Article 106859. https://doi.org/10.1016/j.compbiomed.2023.106859

Peng, W., Lin, J., Dai, W., Chen, G., Fu, X., Liu, L., & Liu, L. (2024). HRLCDR: Hypergraph Representation Learning for Cancer Drug Response Prediction. In W. Peng, Z. Cai, & P. Skums (Eds.), Bioinformatics Research and Applications (Vol. 14955, pp. 339–349). Springer Nature Singapore. doi: 10.1007/978-981-97-5131-0_29.

Peng, W., Liu, H., Dai, W., Yu, N., & Wang, J. (2022). NIHGCN:Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions. *Bioinformatics, 38*(19), 4546–4553. https://doi.org/10.1093/bioinformatics/btac574

Saifuddin, K. M., Bumgardner, B., Tanvir, F., & Akbas, E. (2023). HyGNN: Drug-Drug Interaction Prediction via Hypergraph Neural. *Network, No. arXiv:2206.12747*, arXiv. https://doi.org/10.48550/arXiv.2206.12747

Sen, B., Peng, S., Tang, X., Erickson, H. S., Galindo, H., Mazumdar, T., Stewart, D. J., Wistuba, I., & Johnson, F. M. (2012). Kinase-Impaired BRAF Mutations in Lung Cancer Confer Sensitivity to Dasatinib. *Science Translational Medicine, 4*(136), Article 136ra70-136ra70. https://doi.org/10.1126/scitranslmed.3003513

Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., & Ester, M. (2019). MOLI: Multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics, 35(14), i501–i509. doi: 10.1093/bioinformatics/btz318.

Shehzad, A., Xia, F., Abid, S., Peng, C., Yu, S., Zhang, D., & Verspoor, K. (2024). *Graph Transformers: A Survey, No. arXiv:2407.09777*, arXiv. https://doi.org/10.48550/arXiv.2407.09777

Song, C., Liu, X., Xiong, Z., Liu, L., & Zhang, W. (2025). GNNDRP: Graph Neural Network With Multi-Task Learning for Drug Response Prediction. *IEEE Transactions on Computational Biology and Bioinformatics, 1–9*. https://doi.org/10.1109/TCBBIO.2025.3548692

Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J., Scherf, U., Lee, J. K., Reinhold, W. O., Weinstein, J. N., Mesirov, J. P., Lander, E. S., & Golub, T. R. (2001). Chemosensitivity prediction by transcriptional profiling. *Proceedings of the National Academy of Sciences, 98*(19), 10787–10792. https://doi.org/10.1073/pnas.191368598

Su, R., Liu, X., Wei, L., & Zou, Q. (2019). DeepForest: A deep forest model to predict anti-cancer drug response. *Methods, 166*, 91–102. https://doi.org/10.1016/j.ymeth.2019.02.009

Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., & Wang, F. (2020). Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics, 21*(3), 919–935. https://doi.org/10.1093/bib/bbz042

Tao, W., Liu, Y., Lin, X., Song, B., & Zeng, X. (2023). DGCL: Prediction of multi-relational drug–gene interaction via Dynamic hyperGraph Contrastive Learning. *Briefings in Bioinformatics, 24*(6), Article bbad371. https://doi.org/10.1093/bib/bbad371

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks (No. arXiv:1710.10903). arXiv. doi: 10.48550/arXiv.1710.10903.

Wang, H., Lin, K., Zhang, Q., Shi, J., Song, X., Wu, J., Zhao, C., & He, K. (2024). HyperTMO: A trusted multi-omics integration framework based on hypergraph convolutional network for patient classification. *Bioinformatics, 40*(4), Article btae159. https://doi.org/10.1093/bioinformatics/btae159

Wang, L., Li, X., Zhang, L., & Gao, Q. (2017). SRMF: Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer, 17*(1), 513. https://doi.org/10.1186/s12885-017-3500-5

Wang, Y., Yu, X., Gu, Y., Li, W., Zhu, K., Chen, L., Tang, Y., & Liu, G. (2024). XGraphCDS: An explainable deep learning model for predicting drug sensitivity from gene pathways and chemical structures. *Computers in Biology and Medicine, 168*, Article 107746. https://doi.org/10.1016/j.compbiomed.2023.107746

Xie, G.-B., Yu, J.-R., Lin, Z.-Y., Gu, G.-S., Chen, R.-B., Xu, H.-J., & Liu, Z.-G. (2024). SHGAE: Prediction of miRNA–disease associations based on strengthened hypergraph convolutional autoencoder. *Computational Biology and Chemistry, 108*, Article 107992. https://doi.org/10.1016/j.compbiolchem.2023.107992

Xu, M., Zhu, Z., Zhao, Y., He, K., Huang, Q., & Zhao, Y. (2024). RedCDR: Dual Relation Distillation for Cancer Drug Response Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 21*(5), 1468–1479. https://doi.org/10.1109/TCBB.2024.3404262

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., Ramaswamy, S., Futreal, P. A., Haber, D. A., Stratton, M. R., Benes, C., McDermott, U., & Garnett, M. J. (2012). GDSC: a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Research, 41(D1), D955–D961. doi: 10.1093/nar/gks1111.

Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., & Liu, X. S. (2015). Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLOS Computational Biology, 11*(9), Article e1004498. https://doi.org/10.1371/journal.pcbi.1004498

Zhang, Y., Wei, Z., Yuan, Y., Li, C., & Huang, W. (2024). EquiPocket: An E(3)-Equivariant Geometric Graph Neural Network for Ligand Binding Site Prediction. *Forty-first International Conference on Machine Learning*.

Zhou, P., Wu, Z., Zeng, X., Wen, G., Ma, J., & Zhu, X. (2023). Totally Dynamic Hypergraph. *Neural Networks., 3*, 2476–2483. https://doi.org/10.24963/ijcai.2023/275

Zhu, J., Xia, Y., Qin, T., Zhou, W., Li, H., & Liu, T.-Y. (2021). Dual-view Molecule Pre-training (No. arXiv:2106.10234). arXiv. doi: 10.48550/arXiv.2106.10234.