RedCDR: Dual Relation Distillation for Cancer Drug Response Prediction

Muhao Xu, Zhenfeng Zhu, Yawei Zhao, Kunlun He, Qinghua Huang, and Yao Zhao

***Abstract*—Based on multi-omics data and drug information, predicting the response of cancer cell lines to drugs is a crucial area of research in modern oncology, as it can promote the development of personalized treatments. Despite the promising performance achieved by existing models, most of them overlook the variations among different omics and lack effective integration of multi-omics data. Moreover, the explicit modeling of cell line/drug attribute and cell line-drug association has not been thoroughly investigated in existing approaches. To address these issues, we propose RedCDR, a dual relation distillation model for cancer drug response (CDR) prediction. Specifically, a parallel dual-branch architecture is designed to enable both the independent learning and interactive fusion feasible for cell line/drug attribute and cell line-drug association information. To facilitate the adaptive interacting integration of multi-omics data, the proposed multi-omics encoder introduces the multiple similarity relations between cell lines and takes the importance of different omics data into account. To accomplish knowledge transfer from the two independent attribute and association branches to their fusion, a dual relation distillation mechanism consisting of representation distillation and prediction distillation is presented. Experiments conducted on the GDSC and CCLE datasets show that RedCDR outperforms previous state-of-the-art approaches in CDR prediction.**

***Index Terms*—Cancer drug response (CDR) prediction, graph convolutional networks, knowledge distillation, multi-omics.**

I. Introduction

CANCER is a complex and heterogeneous disease that has emerged as a leading cause of death, seriously threatening human health. A recent study [1] suggests that each cancer patient has a unique genome profile, making it impossible for a single treatment to be effective for every kind of cancer. As a consequence, it holds significant clinical implications to use computational algorithms to predict the response of each patient to drugs based on their genome profile characteristics and provide personalized treatment plans.

Unlike other applications of machine learning, it is difficult to obtain real clinical data on the responses of the same patient to multiple anti-cancer drugs. Therefore, datasets used for drug response prediction mainly rely on preclinical models such as cell lines. A considerable amount of genomic data has been obtained as a result of recent researches like Cancer Cell Line Encyclopedia (CCLE) [2], [3] and Genomics of Drug Sensitivity in Cancer (GDSC) [4], [5], including mutation, copy number, methylation data, as well as hundreds of drug sensitivity and resistance in different cancer cell lines. By utilizing these valuable data, researchers can explore the mechanism of drug response in cancer therapy and develop machine learning methods for predicting cancer drug response (CDR).

Different computational algorithms have been utilized for predicting CDR in cancer cell lines, involving distinct inputs, models, and learning strategies. Initially, some traditional machine learning methods, such as regression [6], [7], support vector machines [8], [9], random forests [10], [11], and matrix factorization [12], [13], were applied to predict drug response. In addition, a network diffusion-based method [14] has been proposed for drug response prediction. It constructs heterogeneous networks between cell lines and drugs and infers unknown responses based on existing associations. In recent years, deep learning has gained widespread applications across various fields due to its ability to learn the inherent complex patterns. To extract latent representations of cell lines and drugs from large amounts of high-dimensional and noisy data, some deep learning models such as deep neural networks (DNNs) [15], [16], [17] and convolutional neural networks (CNNs) [18], [19] have been applied to effectively characterize the nonlinear relationships between them.

Recently, graph neural networks (GNNs) have gained popularity in the field of medicine [20] and bioinformatics [21], [22], [23]. In the case of drugs, their molecular structures can be viewed as graph structures [24], [25]. Meanwhile, for cell lines, drugs, and the associations between them, a corresponding graph can also be established by taking directly them as nodes and edges. In GraphCDR [26], a graph neural network is applied for CDR prediction using the graph constructed by multi-omics data of cell lines, the molecular structure of drugs, and the known cell line-drug responses. In addition, the contrastive learning is

Manuscript received 26 July 2023; revised 18 March 2024; accepted 14 May 2024. Date of publication 22 May 2024; date of current version 9 October 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021ZD0140407, in part by National High Level Hospital Clinical Research Funding under Grant 2022-PUMCH-C-041, in part by Beijing Natural Science Foundation under Grant 7222313, and the in part by the Natural Science Foundation of China under Grant 62302522. *(Corresponding author: Zhenfeng Zhu.)*

Muhao Xu, Zhenfeng Zhu, and Yao Zhao are with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China (e-mail: mhxu1998@bjtu.edu.cn; zhfzhu@bjtu.edu.cn; yzhao@bjtu.edu.cn).

Yawei Zhao and Kunlun He are with the Medical Big Data Research Center, Chinese PLA General Hospital, Beijing 100853, China (e-mail: csyawei.zhao@gmail.com; kunlunhe@plagh.org).

Qinghua Huang is with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: qhhuang@nwpu.edu.cn).

The source code is available at https://github.com/mhxu1998/RedCDR.

Digital Object Identifier 10.1109/TCBB.2024.3404262

adopted to enhance the generalization ability. Considering both the sensitive and resistant responses, BiG-DRP [27] attempts to construct a bipartite graph to leverage the chemical structure of drugs and the underlying relations between drugs and cell lines. Unlike the previous works, NIHGCN [28] designs a neighborhood interaction-based heterogeneous graph convolution network (GCN) for CDR prediction. Creatively constructing a sparse drug cell line pair network, GADRP [29] employs the initial residual and layer attention-based GCN to predict CDR. To make interpretable CDR prediction, SubCDR [30] measures the pairwise interactions between each subcomponent in drug and in cell line, and constructs an interaction network. Despite the promising performance achieved by these existing models, developing effective CDR prediction still encounters the following challenges.

Challenge 1: How to capture the interactive effects between different omics and focus on more important omics data for achieving multi-omics integration? Despite previous studies [17], [25], [26], [31], [32] have taken into consideration the integration of multi-omics data, they are generally limited to simply extracting features from single omics data independently for integrated representations. To realize the interaction of different omics information, the full-kernel matrix and sparse-kernel matrix of the three omics data are adopted in MOFGCN [33] to perform multi-omics fusion in an iterative way. However, as the kernel matrix of cell lines and the one of drugs are used as the representations of themselves respectively, it inevitably causes the serious loss of the original omics information. Moreover, most of the previous works have failed to consider the importance of different omics in CDR prediction, resulting in a lack of individualized knowledge mining for cancer cell lines of different subtypes.

Challenge 2: How to synergistically combine the cell line/drug attribute and cell line-drug association information for comprehensive representation learning? In fact, the attributes of cell lines and drugs, as well as the associations between them, are derived from different sources and are both correlated and complementary. As a usual way for facilitating the information fusion, the attribute learning module and the association learning module are often executed sequentially, forming a cascaded architecture. For instance, some works [26], [33] have taken the attributes and associations as node features and topological structures, respectively, and employed GCNs to fuse information from these two perspectives. However, this approach will result in a loss of the attribute information to a certain extent during information propagation, leading to inferior performance.

Considering the above constraints, we present a dual Relation distillation model for Cancer Drug Response prediction, namely **RedCDR**. A multi-omics encoder with omics-specific representation learning is proposed to achieve adaptive integration. Furthermore, in pursuit of parallel learning and information fusion of the attribute and association, we design a dual-branch architecture along with a dual relation distillation mechanism. Overall, our main contributions can be highlighted as follows:

- We propose a branch-to-fusion dual relation distillation model for cancer drug response prediction, which enables both the independent learning and interactive fusion feasible for cell line/drug attribute and cell line-drug association information.
- To facilitate the adaptive integration of multi-omics data, the multi-omics encoder introduces the multiple similarity relations between cell lines to enhance the omics-specific representation learning and takes the importance of different omics data into account.
- To accomplish knowledge transfer from the two branches to the fusion module, we design a dual relation distillation mechanism consisting of representation distillation and prediction distillation.
- We conduct extensive experiments and analyses on the GDSC and CCLE datasets and illustrate that RedCDR outperforms baselines for CDR prediction.

II. Materials

A. Multi-Omics Data for Cell Lines

Following the data preparation of DeepCDR [21] and GraphCDR [26], we collect three omics data from the CCLE database [2], including genomic mutation data, gene expression data and DNA methylation data.

In particular, we obtain the genomic mutation data, which includes a total of $Dim_g = 34673$ unique mutation positions within the relevant genes (697 genes from COSMIC Cancer Gene Census [34]). Regarding gene expression data, we collect the log-normalized TPM value and represent it as a Dim_t-dimensional ($Dim_t = 697$) feature vector. The DNA methylation data is directly obtained from the processed Bisulfite sequencing data of promoter 1 kb upstream TSS region and is finally represented by a Dim_e-dimensional ($Dim_e = 808$) feature vector.

B. Molecular Graph Data for Drugs

The SMILES strings of drugs are obtained from PubChem database [35], and then can be converted into molecular graphs where the nodes and edges denote chemical atoms and bonds according to the DeepChem library [36]. The attribute of each node in a drug is represented as a Dim_d-dimensional ($Dim_d = 75$) feature vector.

C. Cancer Cell Line-Drug Response Data

Based on the GDSC database [5], we obtain a dataset containing 100572 responses across 561 cell lines and 222 drugs. According to the threshold of each drug provided by the reported maximum screening concentration, we binarize all the IC50 values and acquire 11591 sensitive responses and 88981 resistant responses. For CCLE datebase, the activity area value of responses are downloaded and binarized by setting a threshold (sensitive if the z-score normalized active area > 0.8; otherwise resistant) according to [37]. We finally obtain a dataset with 7307 responses across 317 cell lines and 24 drugs, which includes 1375 sensitive ones and 5932 resistant ones. In the following illustration, the cell line and drug set are denoted as $\mathbb{C} = \{c_i\}_{i=1}^{N_C}$,

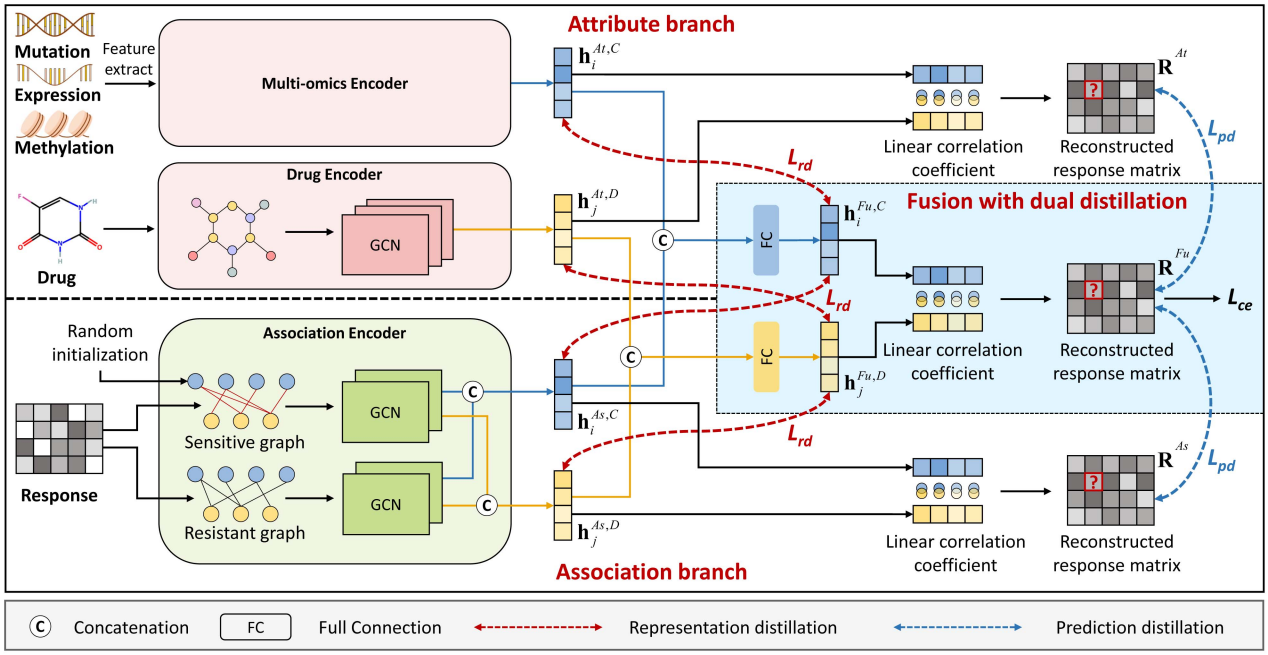

Fig. 1. The framework of the RedCDR model. It consists of three modules: (1) Attribute branch extracts representations from the multi-omics information of cell lines and the molecular structure of drugs; (2) Association branch learns representations from the known sensitive or resistant responses between cell lines and drugs; (3) Fusion module employs dual relation distillation to merge the information from both branches and predict the cell line-drug associations by reconstructing the corresponding matrix.

$\mathbb{D} = \{d_i\}_{i=1}^{N_D}$, where N_C and N_D are the number of each set, respectively.

III. Methods

A. Overview of the Proposed Model

The overall framework of the proposed model is shown in Fig. 1, which mainly consists of three modules:

i) *Attribute branch:* Learns the representations of cell lines by the adaptive interacting multi-omics encoder, and extracts drug attribute representations by the drug encoder from the molecular graphs;

ii) *Association branch:* Divides the known cell line-drug responses into sensitive graph and resistant graph, and learns representations of cell lines and drugs from association information through GCN encoder with randomly initialized node feature;

iii) *Fusion with dual distillation:* Merges representations learned from the two branches and predicts unknown drug responses by calculating linear correlation coefficients between cell lines and drugs. Moreover, the representation distillation and prediction distillation are proposed to facilitate knowledge transfer from the dual branches to the fusion module.

Besides, the necessary notations used in the paper are listed in Table I for ease of understanding.

B. Attribute Branch

1) Adaptive Interacting Multi-Omics Encoder: Previous studies [25], [33] have demonstrated that the drug response prediction can be improved by utilizing multiple omics data simultaneously. Nevertheless, the conventional approach of simple concatenation overlooks the interactive effects between different omics, treating various omics data equally without considering individualized knowledge mining for cancer cell lines of different subtypes. Therefore, to facilitate the integration of multi-omics data and fully consider the correlations and heterogeneity among different omics, we propose an adaptive interacting multi-omics encoder as shown in Fig. 2.

Similar to previous studies [21], [26], different neural networks are first designed to transform raw omics data into latent representations with unified F-dimension as follows:

$$\mathbf{e}_i^g = f_g\left(\mathbf{x}_i^g\right), \quad \mathbf{e}_i^t = f_t\left(\mathbf{x}_i^t\right), \quad \mathbf{e}_i^e = f_e\left(\mathbf{x}_i^e\right) \tag{1}$$

where $\mathbf{x}_i^g \in \mathbb{R}^{1\times Dim_g}$, $\mathbf{x}_i^t \in \mathbb{R}^{1\times Dim_t}$ and $\mathbf{x}_i^e \in \mathbb{R}^{1\times Dim_e}$ are the ith raw cell line data of the genomic, transcriptomic and epigenomic omics. $f_t(\cdot)$ and $f_e(\cdot)$ are fully connected networks to process the transcriptomic and epigenomic data. $f_g(\cdot)$ denotes 1D convolutional network for the genomic data, which is suitable for the linear distribution of mutation positions along the chromosome. Thus, the basic multi-omics representation $\mathbf{h}_i^b \in \mathbb{R}^{1\times F}$ of the ith cell line can be obtained by concatenating all latent representations:

$$\mathbf{h}_i^b = \text{LeakyReLU}\left(\left[\mathbf{e}_i^g || \mathbf{e}_i^t || \mathbf{e}_i^e\right] \mathbf{W}_b + \mathbf{b}_b\right) \tag{2}$$

where $\mathbf{W}_b \in \mathbb{R}^{3F\times F}$ and $\mathbf{b}_b \in \mathbb{R}^{1\times F}$ are the weight matrix and the bias vector, respectively, and $||$ is a vector concatenation operator.

Although the integration of multi-omics data has been achieved, such representations lack feature interaction and importance differentiation among different omics data. Taking into

TABLE I
NOTATIONS USED IN THIS PAPER

Notation	Definition
$m \in \{g, t, e\}$	Different types of omics data, including genomic, transcriptomic and epigenomic omics
$\mathbf{x}_i^m, \mathbf{e}_i^m$	The raw data and latent representation of omics m of the i-th cell line
$\mathbf{h}_i^b$	The basic multi-omics representation of the i-th cell line
$\mathbf{S}_{ij}^m$	The similarity between the i-th and j-th cell line based on omics m
$\mathbf{H}^{m,(l+1)}$	The $(l+1)$-th layer output of GCNs for omics m in the omics-specific representation learning
$\widetilde{\mathbf{A}}^m$	The normalized adjacency matrix for omics m in the omics-specific representation learning
$\mathbf{h}_i^c$	The multi-omics representation of the i-th cell line through dot-product attention
$\mathbf{h}_i^{At,C}, \mathbf{h}_j^{At,D}$	The final representation of the i-th cell line and j-th drug in the attribute branch
$\mathbf{A}^s, \mathbf{A}^r$	The adjacency matrices of the sensitive and resistant graph
$\mathbf{h}_i^{S,C}, \mathbf{h}_j^{S,D}$	The representation of the i-th cell line and j-th drug learned from the sensitive graph
$\mathbf{h}_i^{R,C}, \mathbf{h}_j^{R,D}$	The representations of the i-th cell line and j-th drug learned from the resistant graph
$\mathbf{h}_i^{As,C}, \mathbf{h}_j^{As,D}$	The final representation of the i-th cell line and j-th drug in the association branch
$\mathbf{h}_i^{Fu,C}, \mathbf{h}_j^{Fu,D}$	The final representation of the i-th cell line and j-th drug by the fusion of dual branches
$K \in \{At, As, Fu\}$	Denotes the attribute branch, association branch and the fusion module
$\mathbf{S}_{ij}^{K,C}, \mathbf{S}_{ij}^{K,D}$	The pairwise representation similarity between cell lines and between drugs in the module K
$\hat{\mathbf{R}}_{ij}^K$	The prediction of response between the i-th cell line and j-th drug in the module K

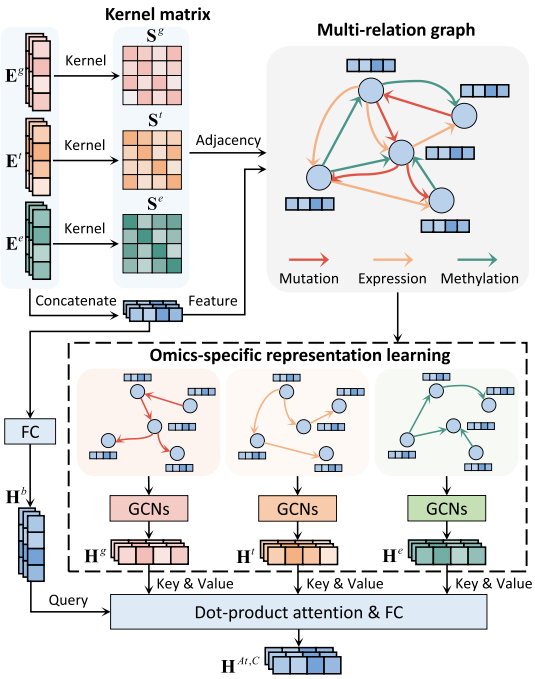

Fig. 2. The structure of adaptive interacting multi-omics encoder.

account the tendency of similar cell lines to exhibit similar representations and drug sensitivities, we introduce the multiple similarity relations among cell lines to enhance the interaction of multi-omics and obtain omics-specific representations. Specifically, the cosine kernel function is uniformly applied to obtain the kernel matrices for each omics, resulting in the generation of a multi-relation graph. Given any cell line pair (c_i, c_j), the kernel matrices of the three omics are defined as:

$$\mathbf{S}_{ij}^m = k\left(\mathbf{e}_i^m, \mathbf{e}_j^m\right) = \frac{\mathbf{e}_i^m \cdot \mathbf{e}_j^{m\mathsf{T}}}{\left\|\mathbf{e}_i^m\right\| \left\|\mathbf{e}_j^m\right\|} \tag{3}$$

where $m \in \{g, t, e\}$ represents different types of omics data. We select the top k similar cell lines for each cell line to construct edges, and finally derive the omics-specific adjacency matrix $\mathbf{A}^m$ for the omics m.

The shared information of similar cell lines propagates through the different omics-specific adjacency relations, facilitating the interaction between different omics and acquiring enhanced omics-specific representations. Specifically, multi-layer GCNs are leveraged to obtain the omics-specific representations of cell lines, and the $(l+1)$th layer output $\mathbf{H}^{m,(l+1)}$ on the m omics is defined as:

$$\mathbf{H}^{m,(l+1)} = \text{LeakyReLU}\left(\widetilde{\mathbf{A}}^m \mathbf{H}^{m,(l)} \mathbf{W}_m^{(l)}\right) \tag{4}$$

where $\mathbf{W}_m^{(l)}$ denotes the transforming mapping matrix of the lth GCN layer for the omics m. The node embedding at the first layer $\mathbf{H}^{m,(0)}$ is initialized as the concatenation of the latent representations of three omics. $\widetilde{\mathbf{A}}^m$ is the normalized adjacency matrix as follows:

$$\begin{aligned} \widetilde{\mathbf{A}}^m &= \left(\hat{\mathbf{D}}^m\right)^{-\frac{1}{2}} \hat{\mathbf{A}}^m \left(\hat{\mathbf{D}}^m\right)^{-\frac{1}{2}} \\ &= \left(\hat{\mathbf{D}}^m\right)^{-\frac{1}{2}} \left(\mathbf{A}^m + \mathbf{I}\right) \left(\hat{\mathbf{D}}^m\right)^{-\frac{1}{2}} \end{aligned} \tag{5}$$

where $\hat{\mathbf{D}}^m$ is the diagonal degree matrix of $\hat{\mathbf{A}}^m$ and $\mathbf{I}$ is the identity matrix.

Due to the heterogeneity of cancer cells, different types of omics play varying degrees of roles in characterizing cell lines. To address this issue, we propose a multi-omics adaptive integration mechanism based on the dot-product attention. By measuring the correlation between the basic representation and each omics-specific representations, the weights assigned to different omics are adaptively adjusted during the integration process. Concretely, the basic multi-omics representation $\mathbf{h}_i^b$ serves as the query $\mathbf{q}_i \in \mathbb{R}^{1\times F}$, and the omics-specific representations $\mathbf{h}_i^m$ serve as the keys $\mathbf{k}_i \in \mathbb{R}^{3\times F}$ and values $\mathbf{v}_i \in \mathbb{R}^{3\times F}$. The multi-omics representation of the ith cell line $\mathbf{h}_i^c$ can be obtained

through dot-product attention:

$$\mathbf{q}_i = \mathbf{h}_i^b, \quad \mathbf{k}_i = \mathbf{v}_i = \left[\mathbf{h}_i^g, \mathbf{h}_i^t, \mathbf{h}_i^e\right] \tag{6}$$

$$\mathbf{h}_i^c = \text{Softmax}\left(\mathbf{q}_i \mathbf{k}_i^{\mathsf{T}}\right) \mathbf{v}_i \tag{7}$$

Ultimately, the final representations of cell lines in the attribute branch $\mathbf{H}^{At,C}$ are denoted as:

$$\mathbf{h}_i^{At,C} = \text{ReLU}\left(\mathbf{h}_i^c \mathbf{W}_C^{At} + \mathbf{b}_C^{At}\right) \tag{8}$$

where $\mathbf{W}_C^{At} \in \mathbb{R}^{F \times F}$ and $\mathbf{b}_C^{At} \in \mathbb{R}^{1 \times F}$ are the weight matrix and the bias vector in the attribute branch, respectively.

2) Drug Encoder: The chemical structure of a drug can be represented as a graph, wherein the nodes and edges correspond to the chemical atoms and bonds, respectively. The molecular graph for drugs can be denoted as $\{\mathcal{G}_j = (\mathbf{X}_j^d, \mathbf{A}_j^d)|_{j=1}^{N_D}\}$, where $\mathbf{X}_j^d \in \mathbb{R}^{N_j^d \times Dim_d}$ and $\mathbf{A}_j^d \in \mathbb{R}^{N_j^d \times N_j^d}$ represent the feature matrix and adjacent matrix of the jth drug, respectively. Here, N_j^d is the number of atoms in the jth drug. Following previous study [26], we apply multi-layer GCNs to capture the latent representation of atom nodes $\mathbf{H}_j^d \in \mathbb{R}^{N_j^d \times F}$. Since different drug molecular graphs have varying numbers of atom nodes, a global max pooling (GMP) layer is utilized to generate a summarized representation of the entire graph as the representation for the jth drug $\mathbf{h}_j^{At,D} \in \mathbb{R}^{1 \times F}$.

C. Association Branch

The associations between cell lines and drugs are established based on their sensitive and resistant responses, which provides complementary insights into the characteristics of both cell lines and drugs beyond their attribute information. For instance, cell lines that demonstrate similar responses to the same drug are expected to possess similar features. Previous studies [28], [33] have typically focused solely on the sensitive responses, while overlooking the valuable information contained in the resistant responses between cell lines and drugs. Although GraphCDR [26] incorporates the joint utilization of two types of responses, it primarily relies on the resistant graph for contrastive learning to assist in the representation learning from the sensitive graph. In response to this problem, the proposed association branch captures the underlying representations of cell lines and drugs in both types of associations simultaneously.

Specifically, the known associations between cell lines and drugs can be constructed as the heterogeneous bipartite graph. The cell line-drug association matrix $\mathbf{R} \in \mathbb{R}^{N_C \times N_D}$ is denoted as:

$$\mathbf{R}_{ij} = \begin{cases} 1, & \text{cell line } c_i \text{ is sensitive to drug } d_i \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Based on the different response types, we build the sensitive and resistant graph correspondingly, which can be denoted as $\mathcal{G}^s = (\mathbf{X}^s, \mathbf{A}^s)$ and $\mathcal{G}^r = (\mathbf{X}^r, \mathbf{A}^r)$. Unlike prior studies, the node embeddings of cell lines and drugs, denoted as $\mathbf{X}^s = \mathbf{X}^r = \begin{bmatrix} \mathbf{X}^C \\ \mathbf{X}^D \end{bmatrix} \in \mathbb{R}^{(N_C+N_D) \times F}$, are trainable randomly initialized vectors, enabling GCN to learn representations from association information independently through neighborhood aggregation. The adjacency matrices of the sensitive and resistant graph are $\mathbf{A}^s \in \mathbb{R}^{(N_C+N_D) \times (N_C+N_D)}$ and $\mathbf{A}^r \in \mathbb{R}^{(N_C+N_D) \times (N_C+N_D)}$, defined as follows:

$$\mathbf{A}^s = \begin{bmatrix} 0 & \mathbf{R}^s \\ (\mathbf{R}^s)^{\mathsf{T}} & 0 \end{bmatrix}, \quad \mathbf{A}^r = \begin{bmatrix} 0 & \mathbf{R}^r \\ (\mathbf{R}^r)^{\mathsf{T}} & 0 \end{bmatrix} \tag{10}$$

where $\mathbf{R}_{ij}^s = \mathbf{M}_{ij}\mathbf{R}_{ij}$ and $\mathbf{R}_{ij}^r = \mathbf{M}_{ij}(1 - \mathbf{R}_{ij})$, which are the sensitive and resistant cell line-drug association matrix, respectively. $\mathbf{M}$ is an indicator matrix, where $\mathbf{M}_{ij} = 1$ indicates the existence of association between cell line c_i and drug d_j, while $\mathbf{M}_{ij} = 0$ indicates the absence of such association.

To obtain association-aware representations, multi-layer GCNs are adopted to the sensitive and resistant graphs. As the identical approach is employed for both the sensitive and resistant graphs, we will present the sensitive graph as an illustrative example below. The $(l+1)$th layer output of the sensitive graph $\mathbf{H}^{s,(l+1)}$ can be represented as:

$$\mathbf{H}^{s,(l+1)} = \text{LeakyReLU}\left(\left(\hat{\mathbf{D}}^s\right)^{-\frac{1}{2}} \mathbf{A}^s \left(\hat{\mathbf{D}}^s\right)^{-\frac{1}{2}} \mathbf{H}^{s,(l)} \mathbf{W}_s^{(l)}\right) \tag{11}$$

where $\hat{\mathbf{D}}^s$ is the diagonal degree matrix of $\mathbf{A}^s$, and the initial input is $\mathbf{H}^{s,(0)} = \mathbf{X}^s$. After several rounds of message propagation, the representations of all cell lines and drugs learned from the sensitive graph can be defined as:

$$\mathbf{H}^{S,C} = \left[\mathbf{h}_0^S, \ldots, \mathbf{h}_{N_C}^S\right], \quad \mathbf{H}^{S,D} = \left[\mathbf{h}_{N_C}^S, \ldots, \mathbf{h}_{N_C+N_D}^S\right] \tag{12}$$

where $\mathbf{h}_i^S = [\mathbf{h}_i^{s,(0)}||\mathbf{h}_i^{s,(1)}||\cdots||\mathbf{h}_i^{s,(L)}]$ represents the concatenation of the outputs from all layers, and L is the number of layers of GCNs.

In the case of the resistant graph, the learned representations of cell lines and drugs, denoted as $\mathbf{H}^{R,C}$ and $\mathbf{H}^{R,D}$ respectively, are calculated in the same approach as that for the sensitive graph. To accomplish the information integration of sensitive and resistant associations, the final representations of cell lines and drugs in the association branch are denoted as:

$$\mathbf{h}_i^{As,C} = \text{ReLU}\left(\left[\mathbf{h}_i^{S,C}||\mathbf{h}_i^{R,C}\right] \mathbf{W}_C^{As} + \mathbf{b}_C^{As}\right) \tag{13}$$

$$\mathbf{h}_j^{As,D} = \text{ReLU}\left(\left[\mathbf{h}_j^{S,D}||\mathbf{h}_j^{R,D}\right] \mathbf{W}_D^{As} + \mathbf{b}_D^{As}\right) \tag{14}$$

where $\mathbf{W}_C^{As}$ and $\mathbf{W}_D^{As}$ are the weight matrices, $\mathbf{b}_C^{As}$ and $\mathbf{b}_D^{As}$ are the bias vectors in the association branch.

D. Fusion With Dual Relation Distillation

The aforementioned dual branches primarily concentrate on the attribute and the association information of cell lines and drugs individually, and the fusion of these two types of information can be achieved through addition or concatenation. However, this approach may not effectively adapt to the differences in the distribution of information between the dual branches, resulting in insufficient preservation of the intra- and inter-relations for cell lines and drugs learned by each branch. Here, the intra-relations refer to the cell line-cell line relations and drug-drug relations, while inter-relations refer to the cell line-drug relations. In order to achieve knowledge transfer from

the dual branches to the fusion module, we propose a dual relation distillation mechanism, which guides the learning of the dual-branch encoders and the final fusion representations.

1) The Fusion of Dual Branches: Concretely, the cell lines and drugs fusion representations of dual branches can be denoted as:

$$\mathbf{h}_i^{Fu,C} = \text{ReLU}\left(\left[\mathbf{h}_i^{At,C}||\mathbf{h}_i^{As,C}\right]\mathbf{W}_C^{Fu} + \mathbf{b}_C^{Fu}\right) \quad (15)$$

$$\mathbf{h}_j^{Fu,D} = \text{ReLU}\left(\left[\mathbf{h}_j^{At,D}||\mathbf{h}_j^{As,D}\right]\mathbf{W}_D^{Fu} + \mathbf{b}_D^{Fu}\right) \quad (16)$$

where $\mathbf{W}_C^{Fu}$ and $\mathbf{W}_D^{Fu}$ are the weight matrices, $\mathbf{b}_C^{Fu}$ and $\mathbf{b}_D^{Fu}$ are the bias vectors.

The linear correlation coefficient is employed as a decoder to reconstruct the cell line-drug association matrix. The definition of the linear correlation coefficient between the two representations is as follows:

$$\rho\left(\mathbf{h}_i, \mathbf{h}_j\right) = \frac{\left(\mathbf{h}_i - \mu_i\mathbf{1}\right)\left(\mathbf{h}_j - \mu_j\mathbf{1}\right)^{\mathsf{T}}}{\sqrt{\left(\mathbf{h}_i - \mu_i\mathbf{1}\right)\left(\mathbf{h}_i - \mu_i\mathbf{1}\right)^{\mathsf{T}}}\sqrt{\left(\mathbf{h}_j - \mu_j\mathbf{1}\right)\left(\mathbf{h}_j - \mu_j\mathbf{1}\right)^{\mathsf{T}}}} \quad (17)$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the representations of the cell line c_i and the drug d_j, respectively. μ_i and μ_j are the mean values of $\mathbf{h}_i$ and $\mathbf{h}_j$, and $\mathbf{1}$ is an all 1 vector with the same dimension as $\mathbf{h}_i$ and $\mathbf{h}_j$. To address the issue of imbalanced positive and negative samples in the response data, we apply the sigmoid activation function with a scaling parameter γ to the output:

$$\phi(x) = \frac{1}{1 + e^{-\gamma x}} \quad (18)$$

Finally, the cell line-drug association matrix from two branches and the fusion module $\hat{\mathbf{R}}_{ij}^K$ can be reconstructed as:

$$\hat{\mathbf{R}}_{ij}^K = \phi\left(\rho\left(\mathbf{h}_i^{K,C}, \mathbf{h}_j^{K,D}\right)\right) \quad (19)$$

where $K \in \{At, As, Fu\}$ denotes the two branches and fusion module respectively.

2) Representation Distillation: The similarity relations between the cell lines and between the drugs in the two branches should be preserved in the fusion representations. In order to transfer the acquired knowledge of intra-relations between the representations of the two branches to the fusion representations, we propose a representation distillation strategy.

Specifically, the cosine kernel function is applied to measure the pairwise representation similarity between the cell lines and between the drugs:

$$\mathbf{S}_{ij}^{K,C} = k\left(\mathbf{h}_i^{K,C}, \mathbf{h}_j^{K,C}\right), \quad \mathbf{S}_{ij}^{K,D} = k\left(\mathbf{h}_i^{K,D}, \mathbf{h}_j^{K,D}\right) \quad (20)$$

To ensure that the inter-relations between the fusion module and the two branches approximates, the Mean Squared Error (MSE) is adopted to quantify the representation distillation loss. As an example, the loss between the attribute branch and fusion module is shown below:

$$\mathcal{L}_{rd}^{At} = \frac{1}{N_C^2}\sum_{c_i\in\mathbb{C}}\sum_{c_j\in\mathbb{C}}||\mathbf{S}_{ij}^{At,C} - \mathbf{S}_{ij}^{Fu,C}||_2^2 + \frac{1}{N_D^2}\sum_{d_i\in\mathbb{D}}\sum_{d_j\in\mathbb{D}}||\mathbf{S}_{ij}^{At,D} - \mathbf{S}_{ij}^{Fu,D}||_2^2 \quad (21)$$

The representation distillation loss between the association branch and the fusion module $\mathcal{L}_{rd}^{As}$ can be calculated the same way. Finally, the total representation distillation loss is defined as:

$$\mathcal{L}_{rd} = \mathcal{L}_{rd}^{At} + \mathcal{L}_{rd}^{As} \quad (22)$$

3) Prediction Distillation: Meanwhile, the interactive cell line-drug relations (i.e., the response) predicted by the two branches should also be passed to the fusion module to guide the learning of the dual encoders and the fusion representations.

Similarly, MSE is utilized as the distillation loss function of predicted responses between the two branches and the fusion module as follows:

$$\mathcal{L}_{pd} = \frac{1}{N_C \times N_D}\sum_{c_i\in\mathbb{C}}\sum_{d_j\in\mathbb{D}}||\hat{\mathbf{R}}_{ij}^{At} - \hat{\mathbf{R}}_{ij}^{Fu}||_2^2 + \frac{1}{N_C \times N_D}\sum_{c_i\in\mathbb{C}}\sum_{d_j\in\mathbb{D}}||\hat{\mathbf{R}}_{ij}^{As} - \hat{\mathbf{R}}_{ij}^{Fu}||_2^2 \quad (23)$$

E. Optimization

Drug response prediction is based on the prediction results of the fusion module, while the predictions from the two branches are used only for the distillation process. The loss of supervised CDR prediction task can be formulated as:

$$\mathcal{L}_{ce} = -\frac{1}{N_M}\sum_{c_i\in\mathbb{C}}\sum_{d_j\in\mathbb{D}}\mathbf{M}_{ij}\left[\mathbf{R}_{ij}log\left(\hat{\mathbf{R}}_{ij}^{Fu}\right) + (1 - \mathbf{R}_{ij})\,log\left(1 - \hat{\mathbf{R}}_{ij}^{Fu}\right)\right] \quad (24)$$

where $\mathbf{M}$ is an indicate matrix and N_M is the sum of all the entries in the matrix. Here, when the association between cell line c_i and drug d_j is in the training set, $\mathbf{M}_{ij} = 1$, otherwise $\mathbf{M}_{ij} = 0$.

The overall objective function combines the prediction task and dual relation distillation tasks simultaneously as follows:

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda_{rd}\mathcal{L}_{rd} + \lambda_{pd}\mathcal{L}_{pd} \quad (25)$$

where λ_{rd} and λ_{pd} are hyperparameters that balance the contributions of different loss functions.

IV. Results and Discussion

A. Baselines and Evaluation Metrics

To assess the effectiveness of our proposed model, we compare it against the following baseline approaches.

- **DeepCDR** [21] is a GCN-based model which integrates multi-omics information of cell lines and chemical structures of drugs.

TABLE II
PERFORMANCE COMPARISON BETWEEN BASELINES AND THE PROPOSED METHOD ON GDSC AND CCLE DATASETS

Dataset	Method	AUC(%)	AUPR(%)	F1(%)	ACC(%)	MCC(%)
GDSC	DeepCDR	82.40 ± 0.43	48.53 ± 0.57	47.46 ± 0.80	87.45 ± 0.59	40.44 ± 0.94
	GraOmicDRP	78.16 ± 0.40	37.91 ± 0.79	41.19 ± 0.58	84.20 ± 0.66	32.73 ± 0.65
	MOFGCN	83.27 ± 0.23	49.45 ± 0.58	48.61 ± 0.73	87.64 ± 0.63	41.70 ± 0.88
	GraphCDR	82.96 ± 0.41	48.69 ± 1.33	48.07 ± 0.73	87.41 ± 0.32	41.00 ± 0.81
	NIHGCN	$\underline{84.33 \pm 0.36}$	$\underline{51.66 \pm 0.86}$	$\underline{50.28 \pm 0.71}$	$\underline{88.01 \pm 0.57}$	$\underline{43.58 \pm 0.81}$
	GADRP	82.20 ± 0.59	47.92 ± 1.32	47.09 ± 1.09	87.43 ± 0.66	40.03 ± 1.40
	RedCDR	$\mathbf{84.67 \pm 0.32}$	$\mathbf{53.07 \pm 0.81}$	$\mathbf{51.20 \pm 0.80}$	$\mathbf{88.73 \pm 0.66}$	$\mathbf{44.90 \pm 1.04}$
	P-value	7.81×10^{-4}	3.12×10^{-6}	4.65×10^{-4}	4.09×10^{-2}	4.79×10^{-4}
CCLE	DeepCDR	95.46 ± 0.56	86.39 ± 1.48	82.24 ± 1.80	93.24 ± 0.67	78.16 ± 2.19
	GraOmicDRP	95.49 ± 0.67	88.30 ± 1.41	82.08 ± 1.60	93.18 ± 0.71	77.99 ± 1.96
	MOFGCN	96.13 ± 0.65	88.93 ± 1.32	82.04 ± 1.82	92.91 ± 0.99	77.76 ± 2.33
	GraphCDR	94.83 ± 1.14	87.46 ± 1.80	82.53 ± 1.66	93.22 ± 0.74	78.40 ± 2.06
	NIHGCN	96.10 ± 0.70	89.25 ± 1.48	$\underline{83.61 \pm 1.65}$	93.57 ± 0.76	$\underline{79.69 \pm 2.05}$
	GADRP	$\underline{96.28 \pm 0.62}$	$\underline{89.94 \pm 1.17}$	83.47 ± 1.53	$\underline{93.65 \pm 0.77}$	79.64 ± 1.90
	RedCDR	$\mathbf{96.62 \pm 0.57}$	$\mathbf{90.42 \pm 1.35}$	$\mathbf{84.55 \pm 1.39}$	$\mathbf{94.03 \pm 0.61}$	$\mathbf{80.90 \pm 1.73}$
	P-value	2.76×10^{-3}	3.93×10^{-2}	2.80×10^{-2}	2.80×10^{-3}	2.84×10^{-2}

The best and second-best results are highlighted in bold and underline, respectively.

- **GraOmicDRP** [25] is a deep learning-based method to integrate molecular graph of drugs and multiple-omics data of cell lines for drug response prediction.
- **MOFGCN** [33] constructs a heterogeneous network that consists of a cell line similarity network, a drug similarity network, and the known drug-cell line associations.
- **GraphCDR** [26] is a graph neural network method with contrastive learning for CDR prediction.
- **NIHGCN** [28] is a heterogeneous graph convolutional network model for CDR prediction based on different levels of neighborhood interaction.
- **GADRP** [29] constructs a sparse drug cell line pairs network and uses the initial residual and layer attention-based GCN to integrate multivariate information from cell lines and drugs.

Our model and all the baselines methods are tested on GDSC and CCLE datasets respectively. In addition, following the previous works [28], [33], we measure the predictive performance by using area under ROC curve (AUC), area under precision/recall curve (AUPR), accuracy (ACC), F1 score (F1) and Matthews correlation coefficient (MCC).

B. Performance Comparison

In order to evaluate the efficacy of our proposed model in predicting cancer drug response, we conducted a comparative analysis with all baseline models. In this experiment, we split all known cell line-drug responses from the GDSC and CCLE datasets randomly into training, validation, and test sets in the ratio of 60%, 20%, and 20%, respectively. To ensure robustness, we repeat the experiment 10 times using different random seeds. We employ Adam with a learning rate from {0.0001, 0.0005, 0.001, 0.005} as the optimizer. The hyperparameters of all comparison baselines adopt the default or optimal settings. For all models, the dimension of cell line and drug representation is 100. For a fair comparison, all the models take the same multi-omics data and drug data as inputs.

Table II shows the superior performance of our model over all baselines on both GDSC and CCLE datasets, in terms of accurately predicting cancer drug responses. The AUC and AUPR values of our model achieved 84.67% and 53.07% on the GDSC dataset, which are 0.34% higher in AUC and 1.41% higher in AUPR compared with the best baseline NIHGCN. On the CCLE dataset, we achieve an AUC of 96.62% and an AUPR of 90.42%, which is an improvement of 0.34% in AUC and 0.48% in AUPR over the best baseline GADRP. The outperformance of our model shows the importance of the adaptive integration of multi-omics data and the independent learning and distillation of both attribute and association information. In addititon, the paired t-test is employed to analyze RedCDR in comparison with the best baseline and the P-value is reported in Table II. It can be seen that the results of RedCDR on different datasets are statistically significant with $P < 0.05$ as tested via the paired t-test. An intuitive observation is that methods constructing networks between cell lines and drugs (i.e., MOFGCN, GraphCDR, NIHGCN, GADRP and our model) tend to perform better than those without construct networks (i.e., DeepCDR and GraOmicDRP). A plausible explanation is that constructing networks can introduce the association information between cell lines and drugs, rather than merely using them as labels. Additionally, the best baseline methods on the two datasets are NIHGCN and GADRP, respectively. Compared to the CCLE dataset, the GDSC dataset has a significantly larger amount of data. GADRP constructs a sparse network of cell line and drug pairs, which facilitates the propagation of information among different cell lines and drugs and contributes to the improved performance. Nevertheless, it first uses an autoencoder for feature extraction, which is not an end-to-end process, and therefore do not perform well on the larger dataset.

In order to evaluate the non-tissue specificity of our model, we group the result according to the tissue types and calculate

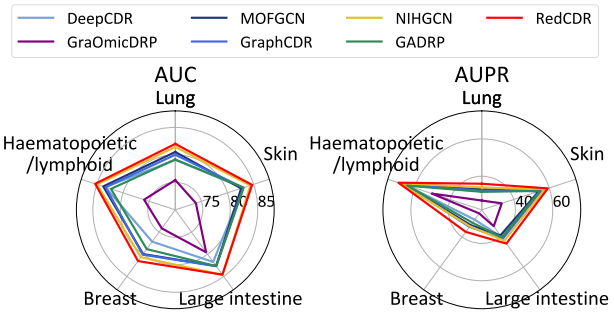

Fig. 3. Results of the five most numerous tissues on the GDSC dataset.

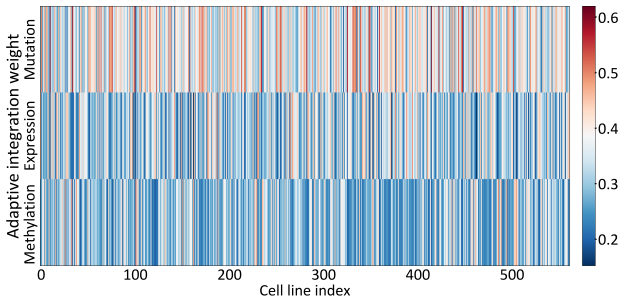

Fig. 4. Adaptive integration weights of multi-omics representations are displayed in a heatmap format, where higher values (or redder colors) indicate greater importance of the omics data.

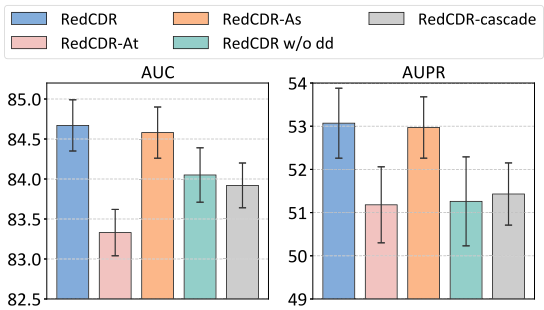

Fig. 5. Results of different variants of two branches.

the AUC and AUPR scores on five tissues that have the largest numbers of cell lines in the GDSC dataset. The five tissues are lung, haematopoietic/lymphoid, breast, large intestine and skin, which account for 58.6% of all cell lines. As shown in Fig. 3, our model performs better than other baseline models on all five tissues.

C. Ablation Study

To evaluate the contributions of the different modules of our model to the prediction performance, we conduct an ablation study by comparing six variants of the model:

- *w/o mg:* Removes the multi-relation graph and omics-specific representation learning, and instead directly employs the basic multi-omics representation.
- *w/o ada:* Removes the multi-omics adaptive integration mechanism and uses the average operation.

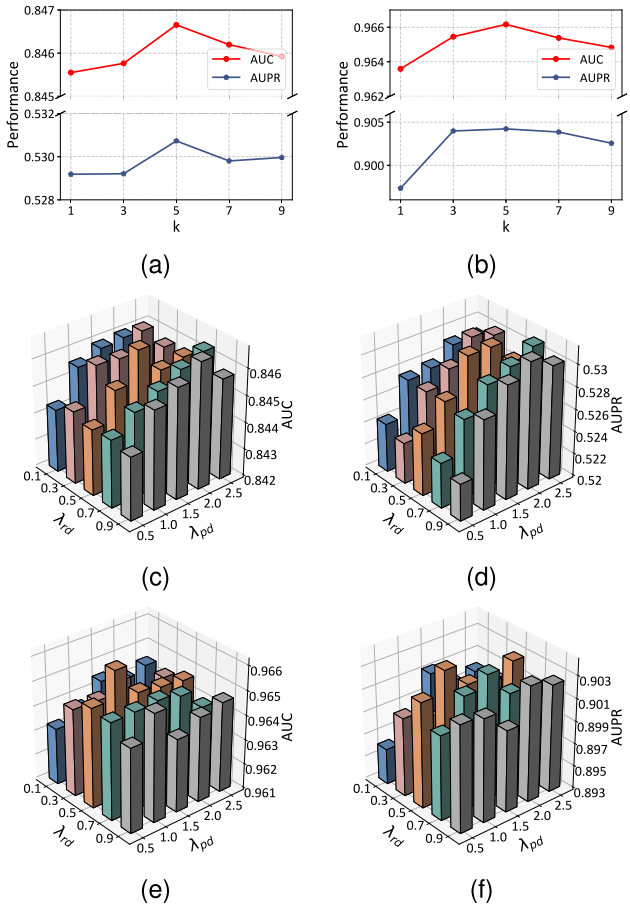

Fig. 6. Hyperparameter sensitivity analysis of RedCDR. (a) and (b) show the performance comparisons equipped with different numbers of similar cell lines on GDSC and CCLE datasets. (c)(d) and (e)(f) show the performance comparisons equipped with different coefficients of dual relation distillation loss on GDSC and CCLE datasets, respectively.

- *w/o sen:* Removes sensitive graph in association branch.
- *w/o res:* Removes resistant graph in association branch.
- *w/o rd:* Removes the representation distillation loss.
- *w/o pd:* Removes the prediction distillation loss.

To further demonstrate the effectiveness of the proposed dual distillation mechanism, we replace it with other fundamental fusion methods:

- *-attn:* Uses a learnable mapping layer to obtain the attention scores of the representations from dual branches, followed by a weighted fusion.
- *-concat:* Directly concatenates the representations from dual branches.
- *-add:* Directly sums representations from dual branches.
- *-mul:* Performs the Hadamard product (i.e., element-wise multiplication) on the representations of dual branches to obtain the fusion representation.

The ablation study is conducted on the GDSC and CCLE datasets. We maintain the same data division and perform 10 repetitions of the experiment as described in the previous subsection. As shown in Table III, the results of the variants (w/o mg) and (w/o ada) show that the importance of introducing the multiple similarity relations between cell lines and the adaptive

TABLE III
ABLATION STUDY ON GDSC AND CCLE DATASETS

Dataset	Method	AUC(%)	AUPR(%)	F1(%)	ACC(%)	MCC(%)
GDSC	w/o mg	84.48 ± 0.29	52.16 ± 0.64	50.63 ± 0.71	88.04 ± 0.50	43.96 ± 0.87
	w/o ada	$\underline{84.58 \pm 0.27}$	$\underline{52.98 \pm 0.92}$	$\underline{50.84 \pm 0.79}$	88.49 ± 0.46	$\underline{44.36 \pm 0.99}$
	w/o sen	84.42 ± 0.30	52.41 ± 0.72	50.69 ± 0.76	$\underline{88.57 \pm 0.29}$	44.25 ± 0.84
	w/o res	84.33 ± 0.26	52.32 ± 0.77	50.54 ± 0.61	88.20 ± 0.58	43.98 ± 0.53
	w/o rd	84.36 ± 0.34	52.57 ± 0.85	50.54 ± 0.88	88.45 ± 0.77	44.15 ± 0.99
	w/o pd	84.32 ± 0.31	51.57 ± 0.78	50.30 ± 0.66	88.23 ± 0.77	43.77 ± 0.97
	-attn	84.15 ± 0.35	51.22 ± 0.84	50.05 ± 0.64	88.13 ± 0.43	43.40 ± 0.73
	-concat	84.05 ± 0.34	51.26 ± 1.03	49.99 ± 0.75	88.11 ± 0.64	43.33 ± 0.95
	-add	84.22 ± 0.33	51.45 ± 0.70	50.10 ± 0.72	88.22 ± 0.62	43.51 ± 0.88
	-mul	84.16 ± 0.26	51.28 ± 0.81	50.13 ± 0.67	87.60 ± 0.33	43.24 ± 0.77
	RedCDR	$\mathbf{84.67 \pm 0.32}$	$\mathbf{53.07 \pm 0.81}$	$\mathbf{51.20 \pm 0.80}$	$\mathbf{88.73 \pm 0.66}$	$\mathbf{44.90 \pm 1.04}$
CCLE	w/o mg	96.30 ± 0.52	89.69 ± 1.24	83.71 ± 1.55	93.78 ± 0.55	79.95 ± 1.77
	w/o ada	$\underline{96.52 \pm 0.56}$	90.22 ± 1.20	$\underline{84.31 \pm 1.20}$	$\mathbf{94.04 \pm 0.47}$	$\underline{80.67 \pm 1.41}$
	w/o sen	96.52 ± 0.58	$\underline{90.27 \pm 1.19}$	83.99 ± 1.39	93.95 ± 0.63	80.36 ± 1.73
	w/o res	96.37 ± 0.57	89.52 ± 1.02	83.49 ± 1.08	93.60 ± 0.51	79.61 ± 1.33
	w/o rd	96.32 ± 0.62	89.70 ± 1.30	83.55 ± 0.87	93.71 ± 0.42	79.69 ± 1.07
	w/o pd	96.36 ± 0.54	89.76 ± 1.12	83.79 ± 1.51	93.82 ± 0.64	80.01 ± 1.85
	-attn	96.40 ± 0.61	89.90 ± 1.39	83.78 ± 1.20	93.78 ± 0.53	79.98 ± 1.50
	-concat	96.19 ± 0.51	89.08 ± 1.18	83.72 ± 1.46	93.77 ± 0.72	79.95 ± 1.84
	-add	96.40 ± 0.59	90.02 ± 1.17	83.62 ± 1.13	93.67 ± 0.52	79.77 ± 1.46
	-mul	96.26 ± 0.61	89.67 ± 1.46	83.10 ± 1.47	93.45 ± 0.62	79.10 ± 1.83
	RedCDR	$\mathbf{96.62 \pm 0.57}$	$\mathbf{90.42 \pm 1.35}$	$\mathbf{84.55 \pm 1.39}$	$\underline{94.03 \pm 0.61}$	$\mathbf{80.90 \pm 1.73}$

The best and second-best results are highlighted in bold and underline, respectively.

integration of multi-omics data. The results of the variants (w/o sen) and (w/o res) suggest that the combination of the two types of association information helps improve the prediction. Consistent with expectations, the performance of the variant models (w/o rd) and (w/o pd) underscores the significant contribution of the dual relation distillation mechanism to the CDR prediction. Comparing several fusion methods, it can be observed that the proposed dual relation distillation mechanism effectively realize feature fusion.

D. Analysis on Model Components and Parameters

1) Analysis on Multi-Omics Adaptive Integration: The weights of each omics data in the multi-omics adaptive integration module on the GDSC dataset are computed and visualized in Fig. 4. The results indicate that the relative importance of each omics data for drug responses varies across different cell lines. It can also be observed from the results that the expression information is slightly less important than the other two omics information on the GDSC dataset.

2) Analysis on the Parallel Dual-Branch: To further investigate the parallel dual-branch architecture and the branch-to-fusion dual relation distillation mechanism, we conduct additional comparative experiments. RedCDR-At and RedCDR-As are models that directly predict CDR using the representations learned from either the attribute branch or the association branch. RedCDR w/o dd removes the dual relation distillation mechanism in the original model. And RedCDR-cascade is a model that cascades the attribute branch and the association branch, applying the representations learned by the attribute branch to the association branch without fusion module and distillation mechanism. As can be seen from Fig. 5, the single branch can also achieve good performance, because the dual relation distillation mechanism enables the two branches to interact with each other and guides the representation learning in each branch. In addition, RedCDR and RedCDR w/o dd performs better than RedCDR-cascade, indicating that the cascade of the attribute branch and the association branch deepens the network and results in a loss of original attribute information to a certain extent during information propagation.

3) Hyperparameter Sensitivity Analysis: The proposed model mainly introduces three hyperparameters, i.e., the numbers of similar cell lines k when performing omics-specific representation learning and the coefficients of dual relation distillation loss $\lambda_{rd}, \lambda_{pd}$. As shown in Fig. 6, we conduct experiments on GDSC and CCLE datasets under different parameter settings. From Fig. 6(a) and (b), it can be observed that a smaller value of k leads to a strong attention to individual similar cell lines, while a larger value introduces irrelevant noise information, both of which fail to achieve optimal performance. As shown in Fig. 6(c-f), we find that the GDSC dataset is more sensitive to λ_{pd}, while the CCLE dataset shows similar sensitivities to both λ_{rd} and λ_{pd}.

4) Computational Complexity Analysis: To investigate the complexity of comparable methods and our proposed model, we present the parameters and average training time per epoch in Table IV. Due to the extensive use of convolutional layers, DeepCDR and GraOmicDRP possess a higher time complexity.

TABLE IV
COMPUTATIONAL COMPLEXITY COMPARISION BETWEEN BASELINES AND THE PROPOSED METHOD

Method	DeepCDR	GraOmicDRP	MOFGCN	GraphCDR	NIHGCN	GADRP	RedCDR
Parameters (M)	0.93	0.53	0.16	0.91	4.63	0.26	1.33
Time/epoch (ms)	1137.32	351.90	5.59	138.08	137.22	177.90	200.86

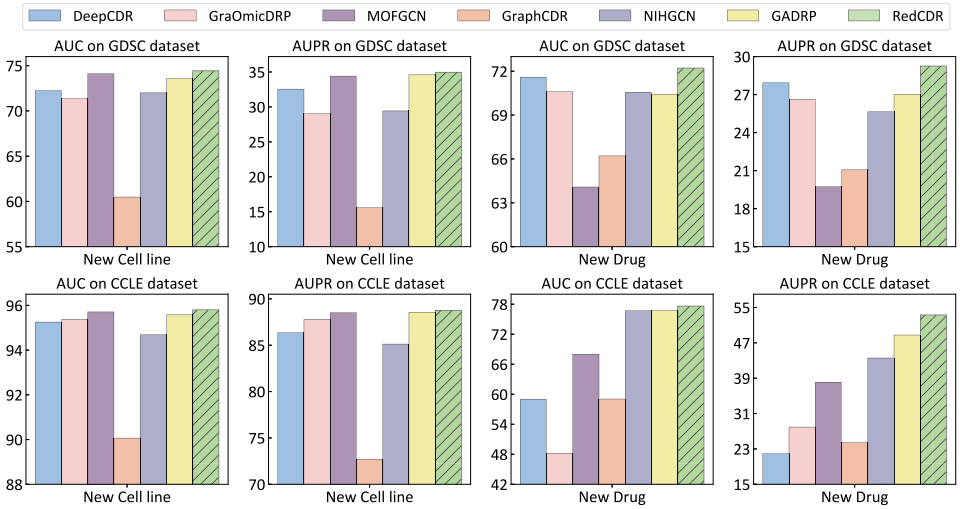

Fig. 7. Results of predicting responses to new cell lines and new drugs on the GDSC and CCLE datasets.

Compared to MOFGCN, GraphCDR, NIHGCN and GADRP, the computational complexity of our model is within acceptable ranges considering the improvement in performance.

E. Predicting Responses to New Cell Lines and New Drugs

With the intention of assessing the inductive capability of our model on new cell lines and drugs, we perform a challenging experiment on the GDSC and CCLE datasets. The dataset is partitioned randomly into five equally sized subsets, where one subset is kept aside as a simulation of new entities in each fold. The prediction model is trained and validated using the remaining four subsets of entities and their corresponding responses.

The results are shown in Fig. 7, where our model achieving the best performance in predicting response to the new cell lines and new drugs. When encountering new cell lines or drugs with no available responses, models relying on the cell line-drug heterogeneity network face significant challenges, as the network lacks relations of these new entities. However, our model can alleviate this issue by the independent attribute and association branches and dual relation distillation mechanism.

F. Regression Experiment

Following the previous work [28], we conduct regression experiment to directly predict the value of IC50 or the z-score normalized active area. We modify the output of the model to predict the response value and evaluate it using three metrics: root mean square error (RMSE), pearson correlation coefficient (PCC), and spearman correlation coefficient (SCC). As shown in the Table V, the experimental results demonstrate the superior performance of our method in predicting the specific response values compared to other baselines.

TABLE V
REGRESSION EXPERIMENT ON GDSC AND CCLE DATASETS

Dataset	Method	RMSE ↓	PCC ↑	SCC ↑
GDSC	DeepCDR	1.108 ± 0.008	0.916 ± 0.002	0.893 ± 0.001
	GraOmicDRP	1.280 ± 0.008	0.886 ± 0.002	0.860 ± 0.002
	MOFGCN	1.160 ± 0.019	0.907 ± 0.003	0.884 ± 0.003
	GraphCDR	1.097 ± 0.007	0.918 ± 0.002	0.895 ± 0.001
	NIHGCN	$\underline{1.081 \pm 0.011}$	$\underline{0.920 \pm 0.002}$	$\underline{0.898 \pm 0.001}$
	GADRP	1.159 ± 0.024	0.908 ± 0.004	0.886 ± 0.005
	RedCDR	$\mathbf{1.047 \pm 0.007}$	$\mathbf{0.925 \pm 0.001}$	$\mathbf{0.904 \pm 0.001}$
	P-value	2.39×10^{-7}	2.39×10^{-7}	1.32×10^{-8}
CCLE	DeepCDR	0.454 ± 0.014	0.892 ± 0.009	0.805 ± 0.013
	GraOmicDRP	0.485 ± 0.011	0.875 ± 0.007	0.781 ± 0.014
	MOFGCN	0.504 ± 0.016	0.864 ± 0.012	0.744 ± 0.016
	GraphCDR	0.498 ± 0.019	0.868 ± 0.012	0.758 ± 0.013
	NIHGCN	0.441 ± 0.015	0.898 ± 0.009	0.812 ± 0.012
	GADRP	$\underline{0.421 \pm 0.012}$	$\underline{0.907 \pm 0.007}$	$\underline{0.819 \pm 0.012}$
	RedCDR	$\mathbf{0.415 \pm 0.012}$	$\mathbf{0.910 \pm 0.006}$	$\mathbf{0.827 \pm 0.012}$
	P-value	1.70×10^{-2}	2.16×10^{-2}	6.45×10^{-4}

The best and second-best results are highlighted in bold and underline, respectively.

G. Case Study of Predicting Unknown Responses

In order to validate the predictive capability of our model for novel cancer cell line-drug responses, we train the proposed model with all known responses in the GDSC dataset, and then predict the unknown responses. By sorting the median predicted

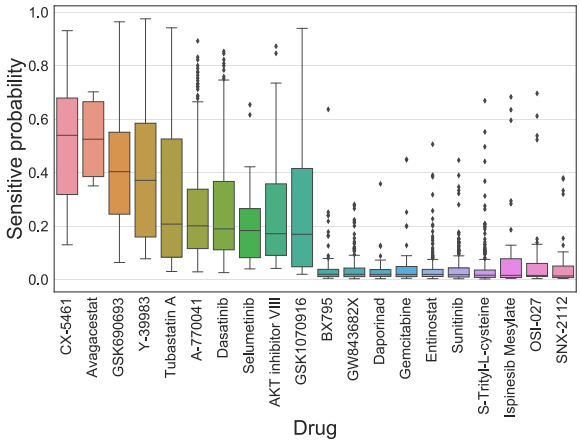

Fig. 8. The predicted sensitive probability of unknown responses that grouped by drugs.

TABLE VI
TOP 5 PREDICTED RESISTANT OR SENSITIVE CANCER CELL LINES TO THE TWO DRUGS

Drug	Prediction	Rank	Cell line	PMID
Dasatinib	Resistant	1	NCI-H1650	N/A
		2	NCI-H196	N/A
		3	**HCC202**	17332353
		4	NCI-H1105	N/A
		5	SW948	N/A
GSK690693	Sensitive	**1**	**RCH-ACV**	19064730
		2	GA-10	N/A
		3	**JeKo-1**	32120074
		4	**MOLT-16**	19064730
		5	MM1-S	N/A

sensitive probability for each drug, Fig. 8 shows the top 10 sensitive and last 10 resistant drugs, respectively.

In light of the unavailability of true values for these unknown responses, we verify the predicted results by referencing relevant literature. Specifically, we focus our analysis on two clinically approved drugs, Dasatinib and GSK690693. Table VI illustrates the top 5 predicted resistant cancer cell lines with Dasatinib and the top 5 predicted sensitive ones with GSK690693. According to the previous stduy [38], the cell line HCC202 could be proved to be resistant to Dasatinib. For the drug GSK690693, three cell lines, RCH-ACV, JeKo-1 and MOLT-16, could be confirmed to be sensitive according to the studies [39], [40]. These case studies show the capability of RedCDR to discover novel cancer cell line-drug responses, thereby facilitating improved drug utilization in clinical trials.

V. CONCLUSION

In this paper, we propose a branch-to-fusion dual relation distillation model for CDR prediction. Specifically, a parallel dual-branch architecture is proposed, which enables both the independent learning and interactive fusion feasible for cell line/drug attribute and cell line-drug association information. To facilitate the adaptive integration of multi-omics data, we design the multi-omics encoder that introduces the multiple similarity relations between cell lines. For the purpose of accomplishing knowledge transfer from the two branches to the fusion module, a dual relation distillation mechanism consisting of representation distillation and prediction distillation is presented. Experiments conducted on the GDSC and CCLE datasets show that the proposed model outperforms previous state-of-the-art approaches in terms of CDR prediction.

Furthermore, it should be pointed out that there are also some limitations for the proposed RedCDR model. First, the current RedCDR utilizes only three types of omics data, and the computational complexity introduced by the multi-relational graph is within an acceptable range. However, it is a consideration worth contemplating how to reduce complexity if more omics data are incorporated in the future. Second, although RedCDR provides a certain degree of interpretability regarding the contribution of different omics data to the final prediction, it fails to capture the specific factors that contribute to sensitivity or resistance to a drug. Finally, how to apply RedCDR to real clinical scenarios through transfer learning to realize personalized patient treatment is also a problem that urgently needs to be addressed.

REFERENCES

[1] M. J. Garnett et al., "Systematic identification of genomic markers of drug sensitivity in cancer cells," *Nature*, vol. 483, no. 7391, pp. 570–575, 2012.
[2] J. Barretina et al., "The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012.
[3] C. C. L. E. Consortium et al., "Pharmacogenomic agreement between two cancer cell line data sets," *Nature*, vol. 528, no. 7580, pp. 84–87, 2015.
[4] W. Yang et al., "Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D955–D961, 2012.
[5] F. Iorio et al., "A landscape of pharmacogenomic interactions in cancer," *Cell*, vol. 166, no. 3, pp. 740–754, 2016.
[6] P. Geeleher, N. J. Cox, and R. S. Huang, "Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines," *Genome Biol.*, vol. 15, pp. 1–12, 2014.
[7] Z. Ding, S. Zu, and J. Gu, "Evaluating the molecule-based prediction of clinical drug responses in cancer," *Bioinformatics*, vol. 32, no. 19, pp. 2891–2895, 2016.
[8] C. Huang, R. Mezencev, J. F. McDonald, and F. Vannberg, "Open source machine-learning algorithms for the prediction of optimal cancer drug therapies," *PLoS One*, vol. 12, no. 10, 2017, Art. no. e0186906.
[9] L. Parca et al., "Modeling cancer drug response through drug-specific informative genes," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 15222.
[10] G. Riddick et al., "Predicting in vitro drug sensitivity using random forests," *Bioinformatics*, vol. 27, no. 2, pp. 220–224, 2011.
[11] Q. Wan and R. Pal, "An ensemble based top performing approach for NCI-dream drug sensitivity prediction challenge," *PLoS One*, vol. 9, no. 6, Art. no. e101183, 2014.
[12] M. Ammad-Ud-Din et al., "Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization," *Bioinformatics*, vol. 32, no. 17, pp. i455–i463, 2016.
[13] L. Wang, X. Li, L. Zhang, and Q. Gao, "Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization," *BMC Cancer*, vol. 17, no. 1, 2017, Art. no. 513.
[14] F. Zhang, M. Wang, J. Xi, J. Yang, and A. Li, "A novel heterogeneous network-based method for drug response prediction in cancer cell lines," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 3355.
[15] L. Rampášek, D. Hidru, P. Smirnov, B. Haibe-Kains, and A. Goldenberg, "Dr VAE: Improving drug response prediction via modeling of drug perturbation effects," *Bioinformatics*, vol. 35, no. 19, pp. 3743–3751, 2019.
[16] M. Li et al., "DeepDSC: A deep learning method to predict drug sensitivity of cancer cell lines," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 2, pp. 575–582, Mar./Apr. 2021.

[17] H. Sharifi-Noghabi, O. Zolotareva, C. C. Collins, and M. Ester, "MOLI: Multi-omics late integration with deep neural networks for drug response prediction," *Bioinformatics*, vol. 35, no. 14, pp. i501–i509, 2019.
[18] Y. Chang et al., "Cancer drug response profile scan (CDRscan): A deep learning model that predicts drug effectiveness from cancer genomic signature," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 8857.
[19] P. Liu, H. Li, S. Li, and K.-S. Leung, "Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network," *BMC Bioinf.*, vol. 20, no. 1, 2019, Art. no. 408.
[20] S. Zheng et al., "Multi-modal graph learning for disease prediction," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2207–2216, Sep. 2022.
[21] 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.
[22] X. Liu, C. Song, S. Liu, M. Li, X. Zhou, and W. Zhang, "Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction," *Bioinformatics*, vol. 38, no. 20, pp. 4782–4789, 2022.
[23] P. Zhang, S. Tu, W. Zhang, and L. Xu, "Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism," *Brief. Bioinf.*, vol. 23, no. 6, 2022, Art. no. bbac 403.
[24] T. Nguyen, G. T. Nguyen, T. Nguyen, and D.-H. Le, "Graph convolutional networks for drug response prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 1, pp. 146–154, Jan./Feb. 2022.
[25] G. T. Nguyen, H. D. Vu, and D.-H. Le, "Integrating molecular graph data of drugs and multiple-omic data of cell lines for drug response prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 2, pp. 710–717, Mar./Apr. 2022.
[26] 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," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab 457.
[27] D. E. Hostallero, Y. Li, and A. Emad, "Looking at the big picture: Incorporating bipartite graphs in drug response prediction," *Bioinformatics*, vol. 38, no. 14, pp. 3609–3620, 2022.
[28] 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.
[29] H. Wang et al., "GADRP: Graph convolutional networks and autoencoders for cancer drug response prediction," *Brief. Bioinf.*, vol. 24, no. 1, 2023, Art. no. bbac 501.
[30] X. Liu and W. Zhang, "A subcomponent-guided deep learning method for interpretable cancer drug response prediction," *PLOS Comput. Biol.*, vol. 19, no. 8, 2023, Art. no. e1011382.
[31] L. Yu, D. Zhou, L. Gao, and Y. Zha, "Prediction of drug response in multilayer networks based on fusion of multiomics data," *Methods*, vol. 192, pp. 85–92, 2021.
[32] Z. Wang, Z. Wang, Y. Huang, L. Lu, and Y. Fu, "A multi-view multi-omics model for cancer drug response prediction," *Appl. Intell.*, vol. 52, no. 13, pp. 14639–14650, 2022.
[33] W. Peng, T. Chen, and W. Dai, "Predicting drug response based on multi-omics fusion and graph convolution," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 1384–1393, Mar. 2022.
[34] Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, and S. A. Forbes, "The cosmic cancer gene census: Describing genetic dysfunction across all human cancers," *Nature Rev. Cancer*, vol. 18, no. 11, pp. 696–705, 2018.
[35] S. Kim et al., "Pubchem 2019 update: Improved access to chemical data," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1102–D1109, 2019.
[36] B. Ramsundar, P. Eastman, P. Walters, V. Pande, K. Leswing, and Z. Wu, *Deep Learning for the Life Sciences*. Sebastopol, CA, USA: O'Reilly Media, 2019.
[37] Z. Dong et al., "Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection," *BMC Cancer*, vol. 15, no. 1, 2015, Art. no. 489.
[38] F. Huang et al., "Identification of candidate molecular markers predicting sensitivity in solid tumors to dasatinib: Rationale for patient selection," *Cancer Res.*, vol. 67, no. 5, pp. 2226–2238, 2007.
[39] D. S. Levy, J. A. Kahana, and R. Kumar, "Akt inhibitor, GSK690693, induces growth inhibition and apoptosis in acute lymphoblastic leukemia cell lines," *Blood, J. Amer. Soc. Hematol.*, vol. 113, no. 8, pp. 1723–1729, 2009.
[40] 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 Chem.*, vol. 97, 2020, Art. no. 103671.

Muhao Xu received the BS degree in school of computers and information technology from Beijing jiaotong University, Beijing, China, in 2020. He is currently working toward the PhD degree in the Institute of Information Science, Beijing Jiaotong University. His research focuses on health informatics and data mining.

Zhenfeng Zhu received the MS degree from the Harbin Institute of Technology, Harbin, China, in 2001 and the PhD degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a professor with the Institute of Information Sciences, Beijing Jiaotong University, Beijing. His current research interests include image and video understanding, computer vision, and machine learning.

Yawei Zhao received the BE, MS, and PhD degrees in computer science from the National University of Defense Technology, China, in 2013, 2015, and 2020, respectively. He is now working with Medical Big Data Research Center of Chinese PLA General Hospital & National Engineering Research Center for the Application Technology of Medical Big Data, Beijing, China. His research interests include federated learning, and medical artificial intelligence.

Kunlun He received the MD degree from the 3rd Military Medical University, Chongqing, China, in 1988, and PhD degree in Cardiology from Chinese PLA Medical School, Beijing, China, in 1999. He worked as a postdoctoral research fellow with the Division of circulatory physiology of Columbia University from 1999 to 2003. He is the director and professor of the Medical Big Data Research Center, Chinese PLA General Hospital, Beijing. His research interests include Big Data and artificial intelligence of cardiovascular disease.

Qinghua Huang received the PhD degree in biomedical engineering from the Hong Kong Polytechnic University, Hong Kong, in 2007. He is currently a full professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include multidimensional ultrasonic imaging, medical image analysis, machine learning for medical data, and intelligent computation for various applications.

Yao Zhao (Fellow, IEEE) received the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), Beijing, China, in 1996. From 2001 to 2002, he was a senior research fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, video analysis and understanding, and artificial intelligence. He was named as a Distinguished Young Scholar by the National Science Foundation of China, in 2010 and was elected as a Chang Jiang Scholar of the Ministry of Education of China, in 2013.