

# DRN-CDR: A cancer drug response prediction model using multi-omics and drug features

K.R. Saranya, E.R. Vimina\*

*Department of Computer Science and IT, School of Computing, Amrita Vishwa Vidyapeetham, Kochi Campus, India*



## ARTICLE INFO

### Keywords:

Cancer drug response prediction  
Multi-omics  
Deep-ResNet  
Graph Convolution Network

## ABSTRACT

Cancer drug response (CDR) prediction is an important area of research that aims to personalize cancer therapy, optimizing treatment plans for maximum effectiveness while minimizing potential negative effects. Despite the advancements in Deep learning techniques, the effective integration of multi-omics data for drug response prediction remains challenging. In this paper, a regression method using Deep ResNet for CDR (DRN-CDR) prediction is proposed. We aim to explore the potential of considering sole cancer genes in drug response prediction. Here the multi-omics data such as gene expressions, mutation data, and methylation data along with the molecular structural information of drugs were integrated to predict the IC50 values of drugs. Drug features are extracted by employing a Uniform Graph Convolution Network, while Cell line features are extracted using a combination of Convolutional Neural Network and Fully Connected Networks. These features are then concatenated and fed into a deep ResNet for the prediction of IC50 values between Drug – Cell line pairs. The proposed method yielded higher Pearson's correlation coefficient ( $r_p$ ) of 0.7938 with lowest Root Mean Squared Error (RMSE) value of 0.92 when compared with similar methods of tCNNS, MOLI, DeepCDR, TGSA, NIHGCN, DeepTTA, GraTransDRP and TSGCNN. Further, when the model is extended to a classification problem to categorize drugs as sensitive or resistant, we achieved AUC and AUPR measures of 0.7623 and 0.7691, respectively. The drugs such as Tivozanib, SNX-2112, CGP-60474, PHA-665752, Foretinib etc., exhibited low median IC50 values and were found to be effective anti-cancer drugs. The case studies with different TCGA cancer types also revealed the effectiveness of SNX-2112, CGP-60474, Foretinib, Cisplatin, Vinblastine etc. This consistent pattern strongly suggests the effectiveness of the model in predicting CDR.

## 1. Introduction

Dynamic diseases like cancer have become a severe threat to human health. The pharmaceutical sector places a high value on developing novel drugs to treat cancer. The objective is to customize therapeutic interventions to increase their efficacy and reduce their negative consequences. Several studies on computational techniques have been undertaken using multi-omics data to expand our knowledge of the molecular origins of cancer. However, due to the heterogeneity in patients, drug predictions become more complex as the response varies among patients. As a result, when developing therapeutic solutions, it is vital to analyse drug reactions. A key component of accelerating the discovery of anti-cancer drugs and comprehending cancer biology is the accurate detection of Cancer Drug Response (CDR). Computational methods for CDR prediction utilize extensive data analysis and machine learning approaches to customize treatments according to individual

patient profiles. Through the integration of various omics data and clinical records, these methods enable personalized medicine, enhancing treatment effectiveness, mitigating adverse effects, and expediting drug development. They have the potential to revolutionize healthcare by refining therapy choices and improving patient outcomes through informed data-centric decisions.

In recent studies of computational methods, the multi-omics data of cell lines, such as the gene expressions, mutation data, methylation data etc., were proven to be good predictors of drug responses. Extensive cancer profiles, including the genomic data (e.g., genomic mutation), transcriptomic data (e.g., gene expression) and epigenomic data (e.g., DNA methylation data) are available in Cancer Cell Line Encyclopedia (CCLE) database (Barretina et al., 2012). To examine the drug response to various cancer cell lines, researchers generally rely on the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al., 2012). Typically, the half-maximal inhibitory concentration (IC50) is used as a

\* Corresponding author.

E-mail address: [vimina.er@gmail.com](mailto:vimina.er@gmail.com) (E.R. Vimina).

standard measure of therapeutic responsiveness across cancer cell lines. It indicates how well a medicine responds against a particular cancer cell from growing. For instance, a low IC<sub>50</sub> value indicates a high drug efficiency and suggests that the medicine is responsive to the cancer cell. The drug response prediction models can be either a binary classification problem (Sensitive or Resistant) or a regression problem (Response as continuous values) (Ding et al., 2016; Ramsundar et al., 2019; Peng et al., 2021).

The majority of the recent computational methods are developed as regression models that predict drug responses as IC<sub>50</sub> values between Drug-Cell line pairs. Studies such as Gao et al. (2019), Rampásek et al. (2019), Wang and Li (2020), Sharifi-Noghabi et al. (2019) address the problem of drug response prediction as both a binary classification as well as regression problem.

The approaches for the drug response prediction can be generally categorized into two: a) Single/multi-omics-based methods b) Drug-structural property augmented methods. The former approach includes the techniques referenced as Bomane et al. (2019), Chang et al. (2018), Chiu et al. (2019), Koras et al. (2020) that utilize single or multi-omics features such as mutation data (genomics), gene expressions (transcriptomics), methylation data, copy number variations etc., of cell lines for predicting drug response. The latter approach includes methods such as Sharifi-Noghabi et al. (2019), Güvenç Paltun et al. (2021), Liu et al. (2019) that incorporate various drug properties as input features in addition to the omics data. These drug properties can be the physicochemical and structural characteristics of individual drugs.

Some of the machine learning methods utilize individual omics data of cancer cell lines, particularly genomic mutations or gene expression. For instance, CDRscan (Chang et al., 2018) uses genetic mutation as cancer cell line feature and drug molecular structural information as drug feature which are loaded into an ensemble CNN model for IC<sub>50</sub> prediction. On the other hand, Twin convolutional neural networks (tCNNS) (Liu et al., 2019) uses genetic mutation as cancer cell features and SMILES sequence as drug features. Other than the methods mentioned above, CaDRReS (Suphavilai et al., 2018) use single-cell line features such as gene expression along with the drug features for CDR prediction using the matrix factorization. Another approach as in Wang et al. (2017) uses gene expressions and drug chemical structures to predict anticancer drug responses of cell lines with the help of a similarity-regularized matrix factorization (SRMF). HIWCF (Zhang et al., 2018), NCFGER (Liu et al., 2018b) and WGRMF (Guan et al., 2019) integrates cell line similarity and drug similarity, as determined by gene expression profiles, chemical structure of the drugs, and similarity in drug responses. Some other methods mentioned below use multi-omics such as gene expression, copy number variation, mutation etc., along with drug features for CDR prediction. ADRML (Moughari et al., 2020) integrates gene expression data, mutation data and copy number variation of cell lines with the drug substructures to construct a bipartite graph connecting drugs and cell lines. It then employs similarity-constrained manifold learning that predicts IC<sub>50</sub> values. Similarly, DSPLMF (Emdadi and Eslahchi, 2020) also uses the same features as in ADRML and computes the probability of the cell lines that are sensitive to drugs by using a logistic matrix factorization approach. MinDrug (Meybodi et al., 2021) finds the best subset of drugs and predicts the IC<sub>50</sub> using Elastic-Net regression.

Recent CDR prediction approaches discussed below widely uses deep-learning models. Literatures like Jiang et al. (2022), Peng et al. (2023) uses single omics data of cancer cell line along with drug features while Liu et al. (2022), Chu et al. (2022) employs convolution neural networks for feature extraction. DeepTTA (Jiang et al., 2022) utilizes transformer and a multi-layer neural network for anti-cancer drug responses prediction with gene expressions and drug SMILES sequences as input. TSGCNN (Peng et al., 2023) uses Graph convolution operation on gene expressions and drug molecular features to diffuse similarity information of drugs and cell lines using a heterogeneous network.

GraTransDRP (Chu et al., 2022) employs Graph Transformer to extract the drug representation and Convolutional neural networks to learn the mutation, methylation, and transcriptomics features, while GraphCDR (Liu et al., 2022) integrates mutation data, gene expression data, methylation data and drug structural features, and uses Graph neural networks for CDR prediction. Some of the recent CDR prediction methods are summarized in Table 1.

Literatures such as Jiang et al. (2022), Chu et al. (2022) etc., uses the features of the entire set of genes in their study while some papers such as Liu et al. (2022), Liu and Zhang (2023), Parca et al. (2019) focused solely on cancer genes. The effect of considering cancer gene features for drug response prediction needs further investigation. Since the main focus of our work is on Cancer drug response, we utilize features derived from cancer genes only during model training.

Despite the advancements in CDR prediction, integrating multi-omics data is still challenging. To address this, we propose an enhanced computational method for CDR prediction by integrating various omics data, drug features and deep learning architectures. Although ResNet models have exhibited notable performance in various fields such as image recognition, natural language processing etc., their potential in drug response prediction is not fully explored. Hence, our objective is to develop an effective approach that integrates multiple features using ResNet architecture, while also exploring the impact of considering sole cancer gene features for CDR.

The following are the major contributions of this work:

- Explores the potential of considering sole Cancer Genes in CDR prediction by integrating multi-omics data such as gene expression, mutation data, and DNA methylation data from Cancer Genes.
- Explores the potential of Deep ResNet model in CDR prediction.
- Case studies with 24 known TCGA cancer types.

The paper is further divided into four sections. The materials and methods used in our work are explained in Section 2. Section 3 explains the validation procedures and results obtained; followed by Section 4 briefing the discussion and Section 5 concludes the paper.

## 2. Materials and methods

Combining and analyzing multidimensional datasets using computational approaches offer several advantages that contribute to a comprehensive and accurate understanding of how individuals will respond to specific drugs. Integration of multi-omics data for drug response prediction effectively reveals the complex connections between genes, proteins, metabolites, and drug reactions. This integrative strategy has the potential to revolutionize personalized medicine and improve patient outcomes through treatments based on individual biological profiles. In this paper, we employ Genomic (Mutation data), transcriptomics (Gene expression data), epigenomics (DNA methylation data) and drug molecular structural information to develop a model for CDR prediction (IC<sub>50</sub> values). Many existing literatures such as Jiang et al. (2022), Chu et al. (2022) etc., uses the features of the entire set of genes in their study while some papers such as Liu and Zhang (2023), Parca et al. (2019), Hermida et al. (2022), Xia et al. (2022), Li et al. (2021) and Huang et al. (2021), consider the features of only cancer genes. In this paper, we use the second approach. Hence, in line with the existing literatures we opted to incorporate the idea of exclusively employing the 697 Cancer gene data (from COSMIC cancer gene census) since these genes have been extensively studied and found to play critical roles in the onset and progression of various cancer types.

### 2.1. Data collection and preprocessing

The study involves extracting Drug features, Cell Line features, and known IC<sub>50</sub> values for drug-cell line pairs from three publicly accessible databases:

**Table 1**

Summary of recent deep-learning based drug response prediction (DRP) methods.

| Sl. no. | Methods  | Inputs   | Description  | Dataset                     |
|---------|--|--|--|-----------------------------|
| 1       | HQNN ( <a href="#">Saginalieva et al., 2023</a> ) (2023) | Genomic features, SMILES format of drugs           | Hybrid Quantum Neural Network - Convolutional, graph convolutional, and deep quantum neural layers are combined with deep quantum computing circuits.  | GDSC<br>PubChem             |
| 2       | TSGCNN ( <a href="#">Peng et al., 2023</a> ) (2023)      | Gene Expression, Drug Fingerprint data             | Graph convolution operation on the feature spaces to diffuse similarity information of drugs and cell lines using a heterogeneous network.   | GDSC<br>CCLE                |
| 3       | DNN-PNN ( <a href="#">Chen et al., 2023</a> ) 2022       | Chemical Structures of drugs, Gene Expression data | Utilizes a deep neural network (DNN) with a gene expression profile as input and a product-based deep neural network (PNN) with drug fingerprint features as input   | CCLE<br>PubChem<br>PaDEL    |
| 4       | DeepITA ( <a href="#">Jiang et al., 2022</a> ) (2022)    | Drug- SMILES format, Gene Expression data          | Utilizes transformer for drug representation learning and a multilayer neural network for transcriptomic data prediction of the anti-cancer drug responses   | PubChem<br>GDSC<br>CCLE     |
| 5       | NIHGCN ( <a href="#">Peng et al., 2022</a> ) (2022)      | Chemical Structures of drugs, Gene Expression data | Utilizing a neighbourhood interaction (NI) layer and a graph convolution network layer (GCN), this approach combines node-level characteristics using graph convolution operations and considers the elemental interactions with neighbouring nodes within the NI layer. | CCLE<br>GDSC<br>PDX<br>TCGA |
| 6       | GraTransDRP ( <a href="#">Chu et al., 2022</a> ) (2022)  | Drug molecular fingerprint, Multi-omics data       | Graph transformer to extract the drug representation and Convolutional neural networks were used to learn the mutation, methylation, and transcriptomics features.   | GDSC<br>CCLE                |
| 7       | GraphCDR ( <a href="#">Liu et al., 2022</a> ) (2022)     | Multi-omics data, drug molecular structures        | A method for CDR prediction using graph neural networks and contrastive learning   | CCLE<br>GDSC<br>TCGA        |
| 8       | TGSA ( <a href="#">Zhu et al., 2022</a> ) 2021           | Multi-omics data, drug molecular structures        | Consists of a Similarity Augmentation (SA) module to combine simple as well as   | CCLE<br>GDSC<br>COSMIC      |

**Table 1 (continued)**

| Sl. no. | Methods   | Inputs   | Description  | Dataset               |
|---------|---|--|--|-----------------------|
| 9       | DeepCDR ( <a href="#">Liu et al., 2020</a> ) (2020)             | Multi-omics data, drug molecular structures            | complex information and Twin Graph neural networks (TGDRP). Uses a hybrid model involving of a uniform graph convolutional network and other subnetworks | CCLE<br>GDSC<br>TCGA  |
| 10      | Network-Based DSP ( <a href="#">Ahmed et al., 2020</a> ) (2020) | Gene Expression  | Network-based model for predicting drug response with the help of gene co-expression network employing graph neural networks.                            | NSCLC<br>GDSC<br>TCGA |
| 11      | tCNNS ( <a href="#">Liu et al., 2019</a> ) (2019)               | SMILES format of drugs, genetic features of Cell-lines | Twin Convolutional Neural Network that uses two convolutional networks to extract features of drugs from (SMILES) format and of cancer cell lines.       | GDSC<br>PubChem       |
| 12      | MOLI ( <a href="#">Sharifi-Noghabi et al., 2019</a> ) 2019      | Gene Expression, Copy number, mutation data            | Multiple feed forward encoding sub-networks, one for each input omics data type.   | GDSC<br>TCGA<br>PDX   |

1. Drug molecular structures of 238 drugs are downloaded from PubChem ([Sharifi-Noghabi et al., 2019](#)), which provides information on the biological activities, chemical structures, and properties of millions of small molecules.
2. The Cell-line features are taken from Cancer Cell Line Encyclopedia (CCLE) ([Barretina et al., 2012](#)), a freely accessible database that provides details on the genomic, pharmacological, and transcriptomic data of 1046 cancer cell lines. CCLE database provides the genomics, transcriptomics and epigenomics profiles of cell lines. There are 11,670 records of cell line-drug studies available in the CCLE database where each record contains the experimental data such as drug target, log(IC50), dose and effective area. We specifically consider the three omics data; genomic mutation data, gene expression profiles and DNA methylation data, for our study. Median imputation is used to fill the missing values in these datasets (either due to absence of data or due to unavailability of data). This study explores the potential of sole cancer genes in CDR prediction as they can influence how cancer cells respond to treatment.
3. The IC50 values (natural log-transformed) of different drugs and cancer cell lines serve as an extensive resource offering details of sensitivity of cancer cell lines to a range of anti-cancer drugs, illustrating the interaction between a drug and a cancer cell line. Both GDSC (Genomics of Drug Sensitivity in Cancer) ([Yang et al., 2012](#)) and CCLE contain IC50 values of drug-cell line pairs. In this work, IC50 values are collected from GDSC2 since it provides standardized protocols for drug sensitivity testing across multiple laboratories, ensuring consistency in data generation and reduction in variability compared to IC50 from CCLE. Moreover, the GDSC2 database contains known IC50 values of 297 drugs while CCLE contains IC50 values of only 24 anti-cancer drugs. The IC50 values of known drug-cell line pairs show minimal variation between the GDSC2 and CCLE datasets.

## 2.2. Preprocessing of data

The steps adopted for preprocessing are as follows:

- ✓ Preprocessing the Drug dataset: The GDSC2 database consists of the details about 297 drugs. However, the PubChem IDs are missing for some of the samples (27 drugs). For this study, the drugs without PubChem IDs are not considered. Additionally, many drugs with distinct GDSC IDs may have the same PubChem IDs due to the varying screening circumstances. In our research, these are considered as a single drug with the same PubChem ID. Finally, 238 drugs are selected for the work. The list of drugs used in the study is described in [Supplementary File-Table 1](#).
- ✓ Preprocessing the list of Cell lines: From the list of 1046 Cancer cell lines obtained from CCLE, those with any missing omics data are not considered. Hence, for the study, only 561 cell lines are considered. The list of cell lines used in the study is described in [Supplementary File-Table 2](#).

Finally, the curated dataset contains 107,446 drug-cell line pairs, constituted by 561 cancer cell lines and 238 drugs. The features of these drugs and cell lines are further used for the study.

## 2.3. Drug and cell line feature representation

Instead of directly taking raw input data and predicting the IC<sub>50</sub> values, the model requires preprocessing steps before training. The features of selected 561 cancer cell lines and 238 drugs are given as inputs to the model. Their feature representations are described below.

### 2.3.1. Drug features

Drugs and compounds have unique chemical structures that can be represented as graphs, with chemical atoms as nodes and bonds between them as edges. The number of atoms varies from 5 to 96. The properties of each atom in a drug are represented as a 75-dimensional feature vector  $f_d$ :

$$f_d = \{f_1, f_2, \dots, f_{75}\} \quad (1)$$

where,  $f_{1,2,\dots,75}$  denotes the chemical and topological features including type, degree, hybridization etc. The structure files of drugs (.MOL files of 238 drugs) are downloaded from the PubChem library. Each row of feature matrix corresponds to the attributes of an atom.

### 2.3.2. Cell line features

The multi-omics profiles of cell lines (561 Cell lines) such as Mutation data, Gene expression data and DNA methylation data of cell lines are extracted from CCLE database. 3751 genes are identified in CCLE ([Jin et al., 2023](#)). Each row in the mutation data corresponds to a cell line and the columns denote the genes. The mutation data for each cell line is viewed as a binary vector, with 1 denoting a modified gene and 0 denoting a non-mutated position. A total of 34,672 unique positions are included in the genomic mutation data and the feature vector ( $f_{mu}$ ) is given in (2):

$$f_{mu} = \{f_1, f_2, \dots, f_{34672}\} \quad (2)$$

For gene expression data, log-transformed and normalized TPM (transcript per million) values of gene expressions are extracted. The gene expressions ( $f_{gx}$ ) are represented as a 697-dimension feature vector given in (3)

$$f_{gx} = \{f_1, f_2, \dots, f_{697}\} \quad (3)$$

corresponding to each Cell-line. The DNA methylation data ( $f_{ml}$ ) is represented as an 808-dimensional feature vector in (4):

$$f_{ml} = \{f_1, f_2, \dots, f_{808}\} \quad (4)$$

for each cell lines. The methylation information is acquired directly from processed Bisulphite sequencing data. To address the effect of any missing values, median value interpolation is implemented. We employ the data of 697 genes that are available in COSMIC Cancer Gene Census. [Table 2](#) contains the list of features and their different combinations considered during the evaluation process.

## 2.4. Proposed methodology

Here, a regression model is proposed for CDR prediction. The input features, which include drug molecular structures and multi-omics data from cancer cell lines (genomic, transcriptomic, and epigenomic data), undergo preprocessing (as mentioned in [Section 2.1](#)) before being fed into the model. Therefore, the model does not function entirely in an end-to-end manner, as it relies on intermediate processing steps before training begins. The layout of this approach is depicted in [Fig. 1](#). Steps in the study are summarized as follows:

1. Drug feature extraction: Used a Uniform Graph Convolution Network (UGCN) to extract the drug feature representations from the drug molecular structures.
2. Cell line feature extraction: Convolution Neural Network (CNN) is used for genomic feature extraction and two Fully Connected Networks (FCN) are used to acquire feature representations of transcriptomics and epigenomic data, respectively.
3. Training and Prediction: The extracted features of drugs and cell lines are concatenated and fed into deep ResNet. The deep ResNet is used to train the model to predict the IC<sub>50</sub> values of drug-cell line pairs.
4. Testing and validation: Evaluation metrics such as Pearson's correlation, Spearman's correlation etc., are used to assess the performance of the proposed method.

ResNet represents a highly potent neural network with the ability to incorporate input residuals into a successive stage. Another advantage is that we are allowed to deepen the network by adding multiple layers. The deeper the network is, the more it learns the features. Considering these advantages, we utilized ResNet in our study.

The following sections describe the comprehensive step-by-step procedure of the proposed method:

### 2.4.1. Drug feature extraction using uniform graph convolution network

A drug can be conceptualized as a graph in which atoms serve as nodes, and the connections between these atoms are represented as edges. We extracted the intrinsic chemical attributes of drugs with the help of a graph convolutional network. The features and adjacency matrix of the drugs are taken as inputs. Graph Convolutional Network (GCN) conducts feature extraction at the level of individual nodes within a single graph, while each drug input encapsulates a distinct and unique graph structure. Hence, we use the idea of a Uniform Graph Convolutional Network (UGCN) as in [Liu et al. \(2020\)](#) to process drugs of variable sizes and structures. Since various drugs encompass differing numbers of atoms (ranging from 5 to 96), the scales of the initial drug graphs can differ. To address this, we use a standardized scale graph

**Table 2**  
Features vectors and their combinations used for Prediction.

|                      |   |
|----------------------|---|
| Gene Expression      | $f_{gx}$  |
| DNA methylation      | $f_{ml}$  |
| Genomic Mutation     | $f_{mu}$  |
| Drug features        | $f_d$   |
| Feature Combinations | $\{f_{gx}, f_{ml}\}$<br>$\{f_{mu}, f_{ml}\}$<br>$\{f_{gx}, f_{mu}\}$<br>$\{f_{gx}, f_{ml}, f_{mu}\}$<br>$\{f_{gx}, f_{ml}, f_{mu}, f_d\}$ |

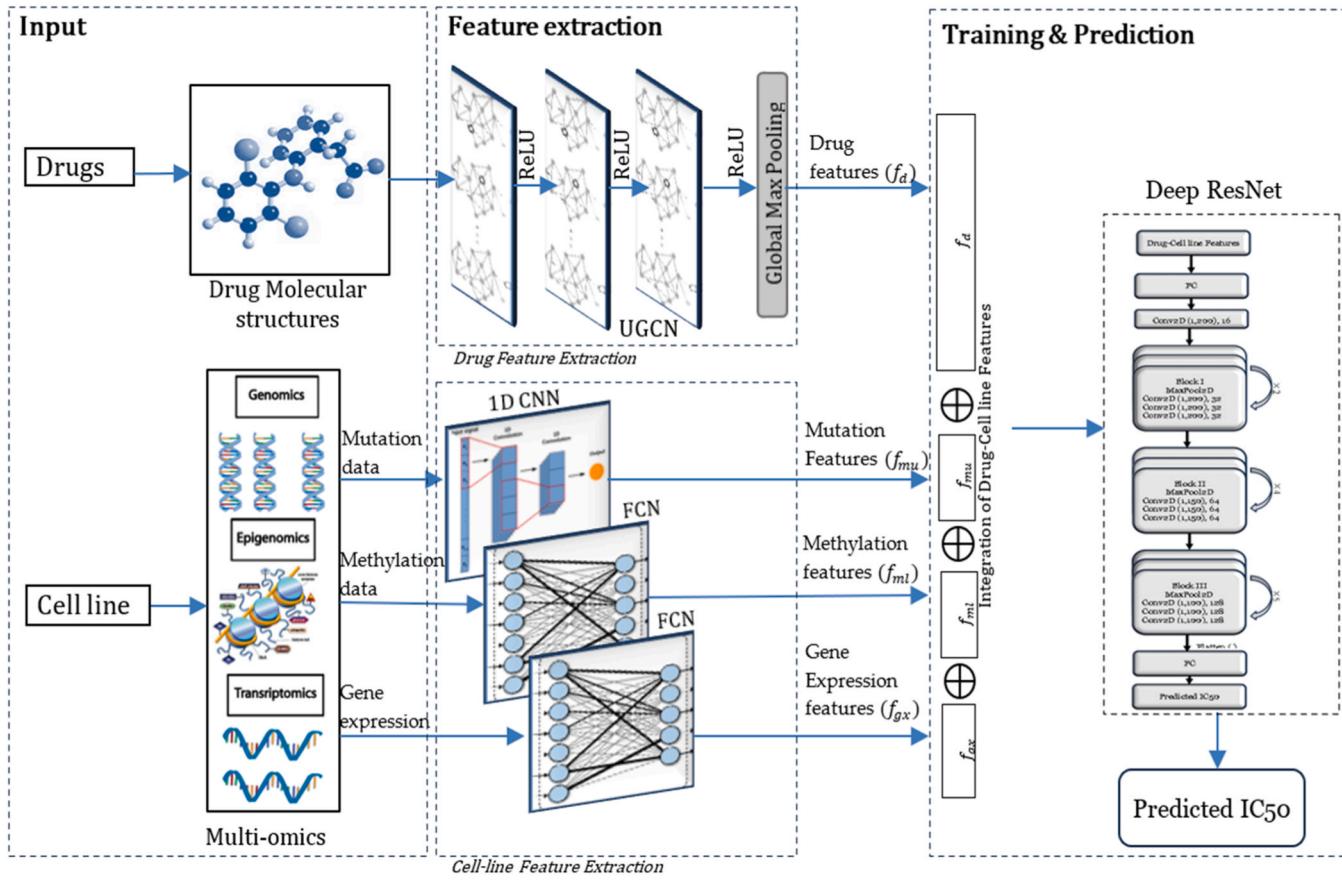


Fig. 1. Model architecture.

onto which we mapped the original graphs. This approach guarantees a consistent application of the GCN across all drugs.

Original graph representation of M drugs:

$$\{\mathcal{G}_i = (X_i, A_i)\}_{i=1}^M \quad (5)$$

complementary graphs are represented as:

$$\{\mathcal{G}_i^c = (X_i^c, A_i^c)\}_{i=1}^M \quad (6)$$

where,  $X_i^c \in \mathbb{R}^{(N-N_i) \times C}$   $A_i^c \in \mathbb{R}^{(N-N_i) \times (N-N_i)}$ , N is a number that is fixed to 100. Hence the standard representation of a drug is given as:

$$A'_i = \begin{bmatrix} A_i & B_i \\ B_i^T & A_i^c \end{bmatrix} \quad (7)$$

$$X_i = \begin{bmatrix} X_i \\ X_i^c \end{bmatrix} \quad (8)$$

where  $A'_i$  and  $X'_i$  are the adjacency matrix and feature matrix of the  $i^{th}$  drug respectively.  $B_i$  represents the conjunction matrix that shows the relationship between the  $i^{th}$  original graph and the complementary graph.

Hence the UGCN applied to the  $i^{th}$  drug can be given as  $f_n(X'_i, A'_i)$  with layer-wise operation as:

$$H_i^{l+1} = \sigma \left( \tilde{D}_i^{-\frac{1}{2}} \tilde{A}'_i \tilde{D}_i^{-\frac{1}{2}} H_i^l \Theta^l \right) \quad (9)$$

where,  $\sigma$  is the activation function,  $\tilde{A}'_i = A'_i + I_N$  is the adjacency matrix

along with self-connections,  $\tilde{D}'_i$  is the degree matrix of  $\tilde{A}'_i$ .  $H_i^l$  and  $\Theta^l$  are the convolved signals and filter parameters of the  $l^{th}$  layer.

The input feature vector of each drug is represented as  $f_d$ . Input to UGCN is a  $238 \times 75$  matrix which denotes 238 drugs of dimension 75.

#### 2.4.2. Multi-omics data integration of cell lines

To integrate multi-omics data, we considered fully connected networks represented as  $y_{gx} = N_{gx}(x_{gx})$ ,  $y_{ml} = N_{ml}(x_{ml})$  that can handle transcriptomics (gene expression data) and epigenomics (methylation data). The genomic data (mutation data) is processed using a CNN denoted as  $y_{mu} = N_{mu}(x_{mu})$ . The networks are mathematically represented as follows:

$$N_{gx} : x_{gx} \in \mathbb{R}^{1 \times d_{gx}} \rightarrow y_{gx} \in \mathbb{R}^{1 \times d} \quad (10)$$

$$N_{ml} : x_{ml} \in \mathbb{R}^{1 \times d_{ml}} \rightarrow y_{ml} \in \mathbb{R}^{1 \times d} \quad (11)$$

$$N_{mu} : x_{mu} \in \mathbb{R}^{1 \times d_{mu}} \rightarrow y_{mu} \in \mathbb{R}^{1 \times d} \quad (12)$$

where,  $d_{gx}$ ,  $d_{mu}$  and  $d_{ml}$  are the latent space dimensions. Due to the binary nature of genomic mutation data and its linear distribution along the chromosome, a 1D Convolutional Neural Network (CNN) was implemented as in (10). Two Dense layer networks are used to handle the transcriptomics and epigenomics and are represented in Eqs. (11) and (12). The input to CNN is the feature vector  $f_{mu}$ . The inputs to the two fully connected networks are  $f_{gx}$  and  $f_{ml}$ . A delayed integration approach was utilized, wherein each network initially acquires a representation of specific omics data within a latent space before being merged.

### 2.4.3. Feature dimensionality reduction

Initially, the features of drugs and cell lines are of different dimensions. Consequently, the individual inputs are transformed into consistent dimensions, allowing us to employ these standardized features for subsequent analysis. We convert these original features individually into a uniform dimension of 50 using Principal component analysis (PCA) after many trials to achieve an ideal balance between dimensionality reduction and information preservation. The transformed features are then integrated together for further steps. The transformed feature vectors are listed below:

$$f_d = \{f_1, f_2, \dots, f_{50}\} \quad (13)$$

$$f_{gx} = \{f_1, f_2, \dots, f_{50}\} \quad (14)$$

$$f_{mu} = \{f_1, f_2, \dots, f_{50}\} \quad (15)$$

$$f_{ml} = \{f_1, f_2, \dots, f_{50}\} \quad (16)$$

Finally, the features are concatenated to form a  $1 \times 200$  drug-cell line feature vector  $f_{cd}$ .

$$f_{cd} = \{f_d, f_{gx}, f_{mu}, f_{ml}\} \quad (17)$$

### 2.4.4. Drug response prediction using ResNet

After dimensionality reduction, the feature representation of drugs and cell lines are concatenated (He et al., 2016) and fed into a deep ResNet for further training of the model. In the proposed method, we use a Deep Residual network with 34 layers which is determined to be the optimal number of layers producing better results after trials. ResNet incorporates skip connections, also referred to as residual connections, which enable the network to skip over specific layers and preserve information from earlier stages of the network. To increase the number of convolutional layers in CNN and deepen the network without encountering the vanishing gradient problem, ResNet is a reliable method.

The residual function  $R(x)$  with input  $x$  and output  $H(x)$  is defined as:

$$R(x) = F(x, \{W_i\}) = W_2 * \delta(W_1 * x) \quad (18)$$

where,  $x$  is the input block,  $F(x, \{W_i\})$  is the residual function approximated by two convolution layers, the learned weights of the two convolution layers and  $\delta$  is the ReLU activation function. The output is calculated as the sum of input  $x$  and the residual  $R(x)$  as given in Eq. (19).

$$H(x) = F(x, \{W_i\}) + x \quad (19)$$

We prefer linear projection shortcuts consisting of a convolutional layer and batch normalization. Deep learning typically involves training a model using a neural network with multiple layers and optimizing using techniques like gradient-based methods. As a result, deep learning demands a lot of computation and the learning is often trapped on a saddle point or local minima. To address this problem, we used the

rectified linear unit (ReLU or  $\delta$ ) as the activation function, whose gradient can be easily determined. ReLU is shown as in (20):

$$\delta(x) = \max(0, x) \quad (20)$$

The combined drugs-cell line features,  $f_{cd}$  is fed to a 1D convolution layer of Resnet architecture with 32 filters and kernel size of  $1 \times 200$ . It is followed by 3 convolution layers of kernel size  $1 \times 200$  and 32 filters all repeated 2 times. Next, we have 3 layers with kernel size  $1 \times 150$  and 64 filters all repeated 4 times. Finally, we have 3 convolution layers with kernel size of  $1 \times 100$  and 128 filters all repeated 5 times. The architecture of the proposed model is given in Fig. 1. Finally, the ResNet generates predictions that correspond to the IC50 values of drug-cell line pairs. Fig. 2 shows the layers of ResNet used in the model.

### 2.5. Model training

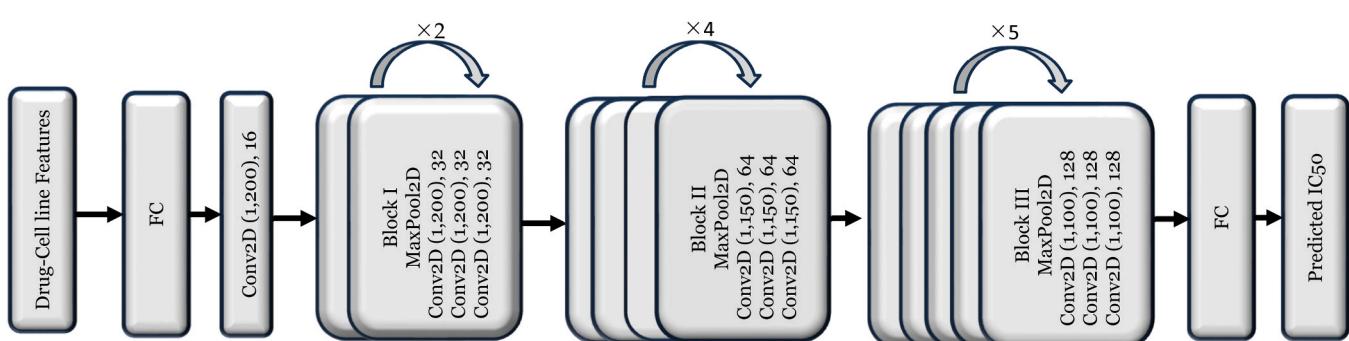
From  $238 \times 561 = 133,518$  instances of drug-cell line interactions, GDSC2 provides IC50 values of 107,446 drug-cell line pairs. Nearly 19.5 % (26,072) of the IC50 values in the GDSC2 database are unknown. 80 % of the known drug-cell line data is used for training the model while the remaining 20 % is used for testing. The split ensures that each of the drug-cell line pairs falls into either training or test set.

Batch normalization and Dropout are used immediately after each convolutional layer to minimize any potential overfitting during the training procedure. For adjusting the parameters during the back-propagation phase, the Adam optimizer is employed. The proposed model uses Root Mean Square Error (RMSE) as the loss function. The number of epochs is fixed at 30. It is feasible to determine an appropriate moment to halt training independently by separating the validation set from the training set, which prevents the issue of overfitting. If the validation RMSE fails to decrease over 5 consecutive epochs, the training is halted using early stopping. Subsequently, the model's predictions on the test set would then be compared with the known IC50 values in GDSC. By changing the hyperparameters of the model through many trials, a progressive increase in the correlation between the known and predicted values of IC50 is observed, resulting in optimal outcomes. Table 3 presents the hyperparameters utilized during model training.

It was discovered that increasing the number of layers and decreasing the pooling size might lead to further performance

**Table 3**  
List of hyper-parameters used in the model.

| Hyperparameters     |       |
|---------------------|-------|
| Epochs              | 30    |
| Learning Rate       | 0.001 |
| Batch Size          | 64    |
| Activation Function | ReLU  |
| Loss                | MSE   |
| Optimizer           | Adam  |
| Dropout Rate        | 0.1   |



**Fig. 2.** ResNet layers.

improvement, although it comes at the cost of reduced stability. This improvement would cause the network to become unstable and diverge during the training. Only those results that assure the stability of the model are presented in this work to maintain the repeatability of the experimental outcomes.

## 2.6. Model evaluation

To assess the performance and quality of the proposed model, regression metrics such as Pearson's ( $r_p$ ) correlation, Spearman's ( $r_s$ ) correlation, RMSE (Root Mean Square Error) and  $R^2$  (Coefficient of Determination) were also employed.  $r_p$  and  $r_s$  correlation coefficients are used to quantify the prediction performance of the model for measuring the correlation between the known and predicted IC50 values of Drug-Cell line pairs. They are computed using the Eqs. (21) and (22), respectively. Linear correlation between the known and predicted IC50 values is measured using  $r_p$ . Unlike  $r_p$ ,  $r_s$  is a non-parametric metric that measures the rank-based correlation between known and predicted IC50 values. The value of  $r_p$  and  $r_s$  ranges between -1 and 1 where a positive correlation indicates a stronger relationship between variables. RMSE is used for finding errors during the training process and computed using Eq. (23). A lower RMSE suggests that the model's predictions are closer to the observed values, indicating better overall performance and accuracy. To evaluate the goodness of fit of the regression model, we used the coefficient of determination ( $R^2$ ). Eq. (24) is used to compute  $R^2$ . It measures how well the regression model explains the variability in the data and is expressed as a value between 0 and 1. A higher value of  $R^2$  is preferable for a better regression model since it explains a greater proportion of the variance in predictions.

$$r_p = \frac{(o_i - \bar{o})(p_i - \bar{p})}{\sqrt{(o_i - \bar{o})^2(p_i - \bar{p})^2}} \quad (21)$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (22)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2} \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (24)$$

where  $n$  is the number of instances or data points,  $o_i$  is the  $i^{th}$  observation,  $p_i$  is the  $i^{th}$  prediction. Mean of  $o_i$  and  $p_i$  are represented as  $\bar{o} = \frac{1}{n} \sum o_i$  and  $\bar{p} = \frac{1}{n} \sum p_i$ . The difference between the two ranks of observation and prediction is denoted as  $d_i = R(o_i) - R(p_i)$ .  $\bar{o}$  represents the mean of the observed values.

The model was also modified to a binary classification problem that categorizes the drugs into sensitive or resistant. The IC50 values binarized using the thresholds provided for each drug as in Iorio et al. (2016), Broussy et al. (2020), Schöffski et al. (2010), Han et al. (2007), Yuan et al. (2021). Binary evaluation metrics such as AUC (Area Under the Curve), AUPR (Area under the Precision-Recall Curve) and ROC-Curve (Receiver Operating Characteristic curve) are used to assess the performance of the binary classification model. Higher values of AUC and AUPR value indicate better model performance, with a maximum value of 1 indicating perfect classification. The AUC and AUPR are calculated using the Eqs. (25) and (26) which give an approximation of the values.

$$AUC \approx \sum FPR[i] - FPR[i-1] \times \frac{TPR[i] + TPR[i-1]}{2} \quad (25)$$

$$AUPR \approx \sum R[i] - R[i-1] \times \frac{P[i] + P[i-1]}{2} \quad (26)$$

where  $FPR$  denotes the False Positive Rate,  $TPR$  denotes the True Positive Rate,  $R$  is the recall and  $P$  is the precision. The ROC curve is a graphical representation of the true positive rate against the false positive rate. In general, a model with a curve that is closer to the top-left corner performs better.

## 3. Results and validation

We evaluated the predictive performance of the model by computing the correlation between predicted and known IC50 values. Additionally, the loss incurred during the model training process is recorded. A comparative assessment of these measures is conducted with similar approaches such as tCNNs, MOLI, DeepCDR, TGSA, NIHGCN, TSGCNN, GraTransDRP and DeepTTA. We employed 5 random cross-validation sets to select the one yielding superior result for further analysis of the model. The Pearson's correlation between the observed and predicted IC50 values in each of the validation sets is depicted in Supplementary File, Fig. 7. It was clear that the proposed method performs better than other methods in different validation sets. The dataset with the highest correlation is chosen for further analysis. The model's proficiency in categorizing drugs as sensitive or resistant is assessed by modifying the model to a binary classifier. The evaluation procedure and results are explained in the following sections.

### 3.1. Performance evaluation of the model using different combinations of features

The combination of drug and cell line features captures the synergistic effects that significantly contribute to drug response prediction. We assessed the model's performance by progressively removing individual features and comparing the results with those obtained when all features were included. This process allowed us to evaluate the contribution of each feature and understand how their absence affects the overall predictive capability of the model. The different combinations of features used in the study are presented in Table 4. The efficacy of the model is assessed using comprehensive metrics, mentioned in Section 2.5. This evaluation process facilitated an understanding of the model's ability to capture relationships and make accurate predictions. It provides valuable insights into the overall performance of the model across different feature sets. The proposed method exhibits better performance when all the features are combined  $\{f_{cd} = f_{gx}, f_{ml}, f_{mu}, f_d\}$ , resulting in the highest correlation. The integration of drug features with cell line features encompassing gene expression data, methylation, and mutation features, exhibited the highest correlation of  $r_p = 0.7938$ ,  $r_s = 0.7804$  and lowest RMSE of 0.9236. This combination demonstrates superior performance when compared to using these features individually. Therefore, we utilize this feature combination for subsequent analysis.

**Table 4**  
Performance analysis of DRN-CDR using different feature combinations.

| Reduced Features                  | Pearson's Correlation Coefficient ( $r_p$ ) | Spearman's Correlation Coefficient ( $r_s$ ) | Root Mean Square Error (RMSE) | Coefficient of Determination ( $R^2$ ) |
|-----------------------------------|---|--|-------------------------------|--|
| $\{f_{gx}, f_{ml}\}$              | 0.7465                                      | 0.7384                                       | 0.9767                        | 0.7202                                 |
| $\{f_{mu}, f_{ml}\}$              | 0.7375                                      | 0.7307                                       | 1.1802                        | 0.7225                                 |
| $\{f_{gx}, f_{mu}\}$              | 0.7433                                      | 0.7316                                       | 0.9823                        | 0.7234                                 |
| $\{f_{gx}, f_{ml}, f_{mu}\}$      | 0.7588                                      | 0.7465                                       | 0.9612                        | 0.7417                                 |
| $\{f_{gx}, f_{ml}, f_{mu}, f_d\}$ | <b>0.7938</b>                               | <b>0.7804</b>                                | <b>0.9236</b>                 | <b>0.7639</b>                          |

### 3.2. Performance analysis of the model using different input feature dimensions

Initially, the features of drugs and cell lines are of different dimensions. Subsequently, Principal Component Analysis (PCA) was utilized to standardize the dimension of input features. The dimensionality of each input feature is reduced to 100, 50, 40, and 10, to evaluate the effect of dimensionality reduction in model prediction. Additionally, we explored non-uniform dimensions for each input. We conducted multiple trials with varying dimensions to ascertain which yields more favourable results. The study revealed that, using a reduced dimension of 50 yielded better results when compared to other dimensions. The details of the analysis are provided in [Supplementary File, Table 3](#).

### 3.3. Evaluation of drug-cell line IC50 values

The model was trained using known drug-cell line pairs, enabling it to subsequently estimate IC50 values for unfamiliar pairs within GDSC. In [Fig. 3](#), the drugs are grouped based on the median of their estimated IC50 values for cell lines across all the cancer types. Drugs exhibiting the highest median IC50 values are potentially ineffective whereas drugs with the lowest median IC50 values indicate their relative effectiveness. High median IC50 values also show a lower affinity of the drug for its target. Drugs with high median IC50 values are generally less potent and may require higher concentrations to achieve the desired inhibitory effect. The drugs such as Tivozanib, SNX-2112 and CGP-60474 were found to be the most effective with the lowest median value. Clinical literatures such as [Zhidkova et al. \(2022\)](#), [Taguchi \(2019\)](#), [Eskens et al. \(2011\)](#), [Santoni et al. \(2018\)](#) with supporting evidence represents CGP-60474 and Tivozanib as promising drug candidates targeting multiple cancers. The other drugs such as PHA-665752, Foretinib, Cisplatin etc., were also found to be more sensitive when compared with the rest. The medical researches such as [Wang et al. \(2021\)](#), [Lee et al. \(2024\)](#), [Liu et al. \(2018a\)](#), [Grojean et al. \(2021\)](#), [Rodler et al. \(2023\)](#) and

[Papadakos et al. \(2024\)](#) highlights the role of PHA-665752, Foretinib and cisplatin in cancer therapy. [Table 5](#) describes the drugs arranged according to the median of estimated log (IC50) values obtained using the proposed method. External evidence in medical literature proved the effectiveness of the above drugs in treating cancers. [Supplementary File - Fig. 1](#) shows the resultant drugs grouped based on the median of their estimated IC50 values when the feature dimensions were reduced to 100.

### 3.4. Comparison with similar methods

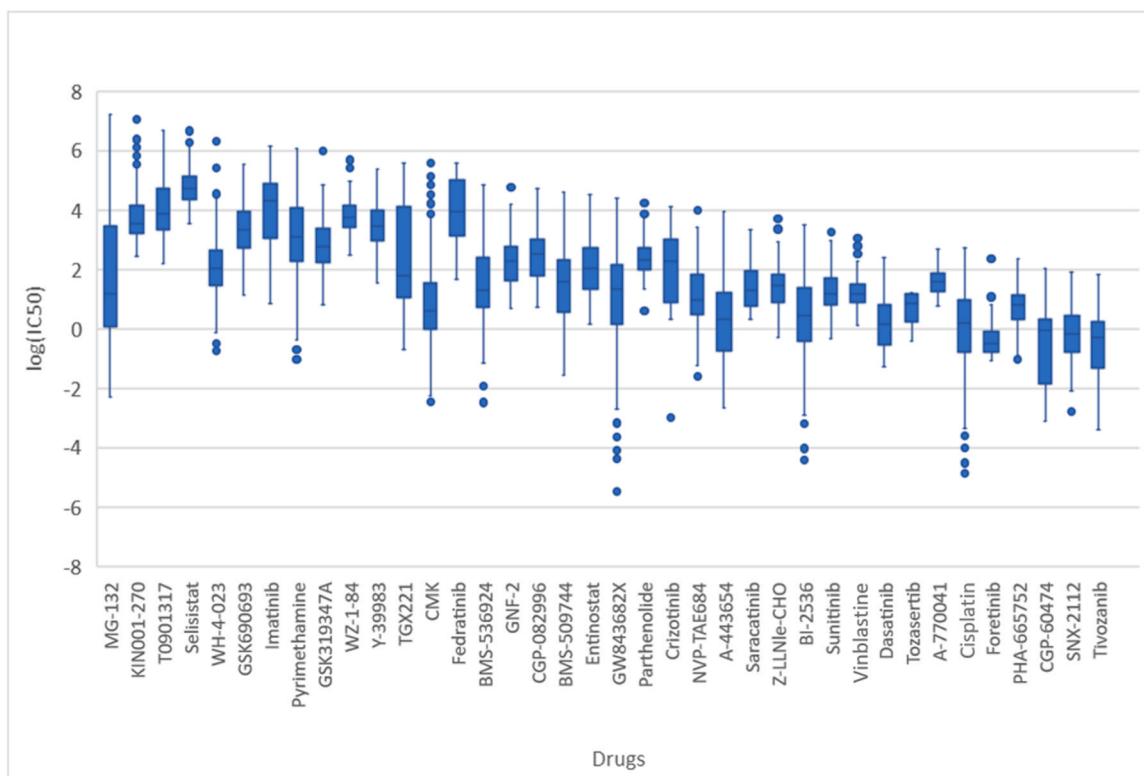
The following scenarios were used for the comparative analysis of the proposed method with similar approaches:

- Assessing the Predictive Power: Correlation and RMSE analysis.
- Model Evaluation using Multiple Datasets
- Best and Worst Predictions: Finding the drugs with best and worst predictions.
- Binary Classification analysis: Extending the model to classify the drugs as sensitive or resistant.

Similar methods considered for conducting a comparative analysis

**Table 5**  
Highly sensitive anti-cancer drugs obtained using the proposed method.

| Sl. No. | Drugs      | Evidence   |
|---------|------------|--|
| 1       | Tivozanib  | <a href="#">Eskens et al. (2011)</a> , <a href="#">Santoni et al. (2018)</a>                                   |
| 2       | SNX-2112   | <a href="#">Kataoka et al. (2012)</a> , <a href="#">Liu et al. (2015)</a> , <a href="#">Wang et al. (2015)</a> |
| 3       | CGP-60474  | <a href="#">Zhidkova et al. (2022)</a> , <a href="#">Taguchi (2019)</a>  |
| 4       | PHA-665752 | <a href="#">Lee et al. (2024)</a> , <a href="#">Yadav et al. (2024)</a>  |
| 5       | Foretinib  | <a href="#">Wang et al. (2021)</a> , <a href="#">Liu et al. (2018a)</a>  |
| 6       | Cisplatin  | <a href="#">Rodler et al. (2023)</a> , <a href="#">Papadakos et al. (2024)</a>                                 |



**Fig. 3.** The predicted IC50 values categorized based on drugs (Reduced Input dimension of 50). The drugs with average predicted IC50 value across the missing cell lines are plotted.

are as follows:

- tCNNS ([Liu et al., 2019](#))- Twin Convolutional Neural Network uses two convolutional networks to extract features of drugs (from SMILES format) and cancer cell lines. The interaction between the drugs and cell lines are predicted with the help of a fully connected network.
- DeepCDR ([Liu et al., 2020](#)) represents a hybrid anti-cancer drug response prediction model that utilizes a GCN-based framework, incorporating multi-omics profiles from both cell lines and drugs.
- TGSA ([Zhu et al., 2022](#)) is one of the methods for drug response prediction using a twin graph neural network with a similarity Augmentation module.
- NIHGCN ([Peng et al., 2022](#)) is a Neighbourhood Interaction-based Heterogeneous Graph Convolution Network method for drug response prediction.
- DeepTTA ([Jiang et al., 2022](#)) uses transformer for drug representation learning and a multilayer neural network for transcriptomic data prediction of the anti-cancer drug responses.
- TSGCNN ([Peng et al., 2023](#)) is based on a two-space graph convolutional neural network, that uses the graph convolution operation on the feature space to diffuse similarity information.
- GraTransDRP ([Chu et al., 2022](#)) is a transformer-based drug response prediction model.
- MOLI ([Sharifi-Noghabi et al., 2019](#)) is a multi-omics late integration method based on deep neural networks.

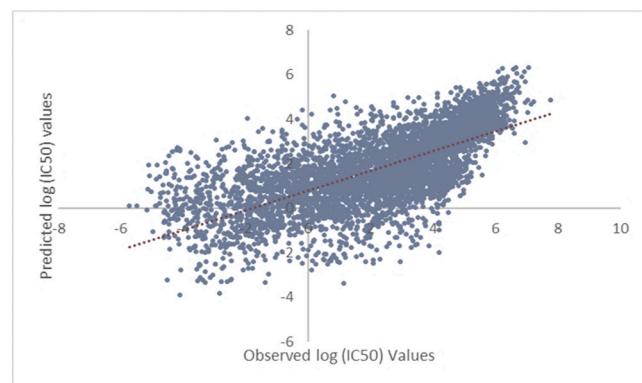
We utilized the same train and test datasets for performance comparison i.e., the dataset consisting of 107,446 known drug-cell line pairs, with 80 % of the data for training and the remaining 20 % for testing. For each of the methods, the features outlined in the original paper were utilized.

#### 3.4.1. Assessing the predictive power

To assess the effectiveness and reliability of the proposed model, a comparative analysis was conducted using the regression metrics mentioned in [Section 2.5](#). The findings from the comparative analysis of recent deep-learning methods such as transformers, graph neural networks etc. These includes tCNN, DeepCDR, TGSA, NIHGCN, TSGCNN, GraTransDRP, MOLI and DeepTTA are provided in [Table 6](#). To ensure the consistency in experimental settings across different methods, we employed identical train and test data.

- The proposed method obtained the lowest Root Mean Squared Error (RMSE) of 0.9236 when compared to other methods, which indicates that the model's predictions on average are closer to the actual IC50 values. Compared methods mentioned above showed RMSE of 1.635, 1.466, 1.434, 1.278, 0.9635, 0.9751, 0.9383, 1.6128 and 0.9203, respectively. This observation emphasizes the model's accuracy in capturing the variability in prediction.
- The relatively highest Pearson's correlation coefficient ( $r_p = 0.7938$ ) and Spearman's correlation coefficient ( $r_s = 0.7804$ ) further supports the effectiveness of the proposed method, revealing a strong linear relationship between the predicted and actual values. Other compared methods showed  $r_p$  values of 0.65422, 0.61736, 0.68334, 0.68923, 0.72762, 0.76841, 0.79265, 0.76275 and 0.72943, respectively.
- Additionally, the higher value of the coefficient of determination  $R^2 = 0.7639$ , implies that the model fits the data well and has the strong predictive capability. The  $R^2$  value obtained by other methods varies between 0.588 and 0.762.

The outcomes of the current model are highly promising, indicating a significant enhancement in its predictive capabilities compared to similar methods. [Fig. 4](#) shows the regression results on the test dataset when compared to the known IC50 values. [Fig. 5](#) depicts the loss (RMSE) over epochs during the training of the proposed model which can provide valuable insights into how well the model is learning. A decreasing trend in RMSE of our model suggests that the model is improving its



**Fig. 4.** Scatter Plot showing the Regression results on test set comparing the Predicted log (IC50) with Observed log (IC50) values.

**Table 6**  
Comparative analysis of similar baseline approaches.

| Method   | Input Features             |                 | $r_p$         | $r_s$         | RMSE          | $R^2$         |
|--|----------------------------|-----------------|---------------|---------------|---------------|---------------|
|  | Cell-Line                  | Drug            |               |               |               |               |
| tCNNS ( <a href="#">Liu et al., 2019</a> ) (2019)          | Sequence (multi-omics)     | Molecular Graph | 0.65422       | 0.64223       | 1.635         | 0.6517        |
| MOLI ( <a href="#">Sharifi-Noghabi et al., 2019</a> ) 2019 | Sequence (multi-omics)     | Molecular Graph | 0.61736       | 0.61014       | 1.6128        | 0.5882        |
| DeepCDR ( <a href="#">Liu et al., 2020</a> ) (2020)        | Sequence (multi-omics)     | Molecular Graph | 0.68334       | 0.66084       | 1.466         | 0.6678        |
| TGSA ( <a href="#">Zhu et al., 2022</a> ) 2021             | Graph (PPI)                | Molecular Graph | 0.68923       | 0.65356       | 1.434         | 0.6540        |
| NIHGCN ( <a href="#">Peng et al., 2022</a> ) (2022)        | Graph (PPI)                | Molecular Graph | 0.72762       | 0.71535       | 1.278         | 0.7223        |
| DeepTTA ( <a href="#">Jiang et al., 2022</a> ) (2022)      | Sequence (Transcriptomics) | SMILES          | 0.76841       | 0.74504       | 0.9635        | 0.7538        |
| GraTransDRP ( <a href="#">Chu et al., 2022</a> ) (2022)    | Sequence (Transcriptomics) | SMILES          | 0.76275       | 0.76133       | 0.9383        | 0.7582        |
| TSGCNN ( <a href="#">Peng et al., 2023</a> ) (2023)        | Sequence (multi-omics)     | Molecular Graph | 0.72943       | 0.70438       | 0.9751        | 0.6994        |
| DRN-CDR  | Sequence (multi-omics)     | Molecular Graph | <b>0.7938</b> | <b>0.7804</b> | <b>0.9236</b> | <b>0.7639</b> |

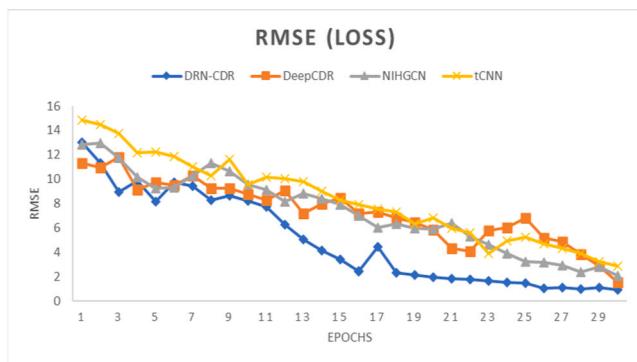


Fig. 5. RMSE or loss reported for each epoch during the training.

prediction accuracy over time.

Further, we have compared the results of the proposed method with some of the non-deep learning methods such as CaDRReS (Suphavilai et al., 2018), ADRML (Moughari et al., 2020) and DSPLMF (Emdadi and Eslahchi, 2020). The results are included in Supplementary file, Table 4. It was observed that Pearson's correlation obtained using the deep-learning approaches is higher when compared to that of the non-deep-learning approaches.

#### 3.4.2. Model evaluation using multiple datasets

To evaluate the predictive power and generalizability of the proposed method across different datasets, we employed two reliable Cell-Line databases such as CCLE and NCI-60 other than GDSC. These datasets contain known IC-50 response values of Drug-Cell line pairs necessary to access the predictive capability of DRN-CDR. We extracted the test datasets from these databases and then applied the pre-trained model to conduct a performance analysis of the proposed method on different datasets. The correlation between observed and predicted response values were computed to access the reliability of the proposed model across different datasets. Details of the comparative analysis conducted using similar approaches is given in Supplementary File – Table 8. Notably, DRN-CDR exhibited highest correlation of  $r_p = 0.7685$ ,  $r_s = 0.7133$  while  $r_p$  ranges from 0.5934 to 0.7621 for other compared methods in the case of CCLE datasets. GraTransDRP performed better on NCI-60 datasets with correlation of  $r_p = 0.7929$  and  $r_s = 0.7372$ . However, DRN-CDR also showed higher correlation of  $r_p = 0.7885$ ,  $r_s = 0.7333$  among the rest of the methods. This observation suggests that the proposed method can exhibit higher predictive performance across different datasets. This ensures the generalizability of the model.

#### 3.4.3. Best and worst predictions

To assess the efficacy of the proposed method, we analysed the correlation values of drugs with best and worst predictions of IC50 among different methods. The results of the comparative analysis indicate that the drugs named Tivozanib, SNX-2112, Cisplatin, Foretinib and Vinblastine from our method exhibit high performance with Pearson's correlation ( $r_p$ ) values of 0.8452, 0.8402, 0.8327, 0.8277 and 0.8231, respectively. The corresponding scatter plots of predictions are given in Fig. 6a–e. Medical literatures such as Eskens et al. (2011), Santoni et al. (2018), Liu et al. (2015), Wang et al. (2015) shows the influence Tivozanib and SNX-2112 in cancer treatment. Other supporting literatures Romani (2022), Tchounwou et al. (2021), Dhyani et al. (2022), Haque et al. (2018), Sohn et al. (2020) and Nazari et al. (2024) reveal the role of Cisplatin, Foretinib and Vinblastine in cancer therapy. The worst-case performance is observed with a drug named Pyrimethamine exhibiting a correlation of  $r_p = 0.2817$ ,  $r_s = 0.3023$ . The corresponding scatter plot is depicted in Fig. 6f.

The Supplementary File also includes comparable outcomes achieved by our method with an input feature dimensionality of 100. The drugs named SNX-2112 and GFN-2 exhibited the best performance

among the rest of the drugs. Moreover, among all drugs examined, SNX-2112 consistently demonstrates the highest correlation, indicating its strong predictive potential in the context of the study. The ability of the drug named SNX-2112 to treat cancer is currently a subject of investigation in many medical researches like Kataoka et al. (2012), Liu et al. (2015), Wang et al. (2015). The scatter plot of SNX-2112 with the highest correlation measures of  $r_p = 0.8176$ ,  $r_s = 0.7986$  respectively which is shown in Supplementary File – Fig. 2. Conversely, the drug Pyrimethamine shows the worst prediction with the lowest correlation measures ( $r_p = 0.3241$ ,  $r_s = 0.3277$ ). The corresponding scatter plot is depicted in Supplementary File – Fig. 3.

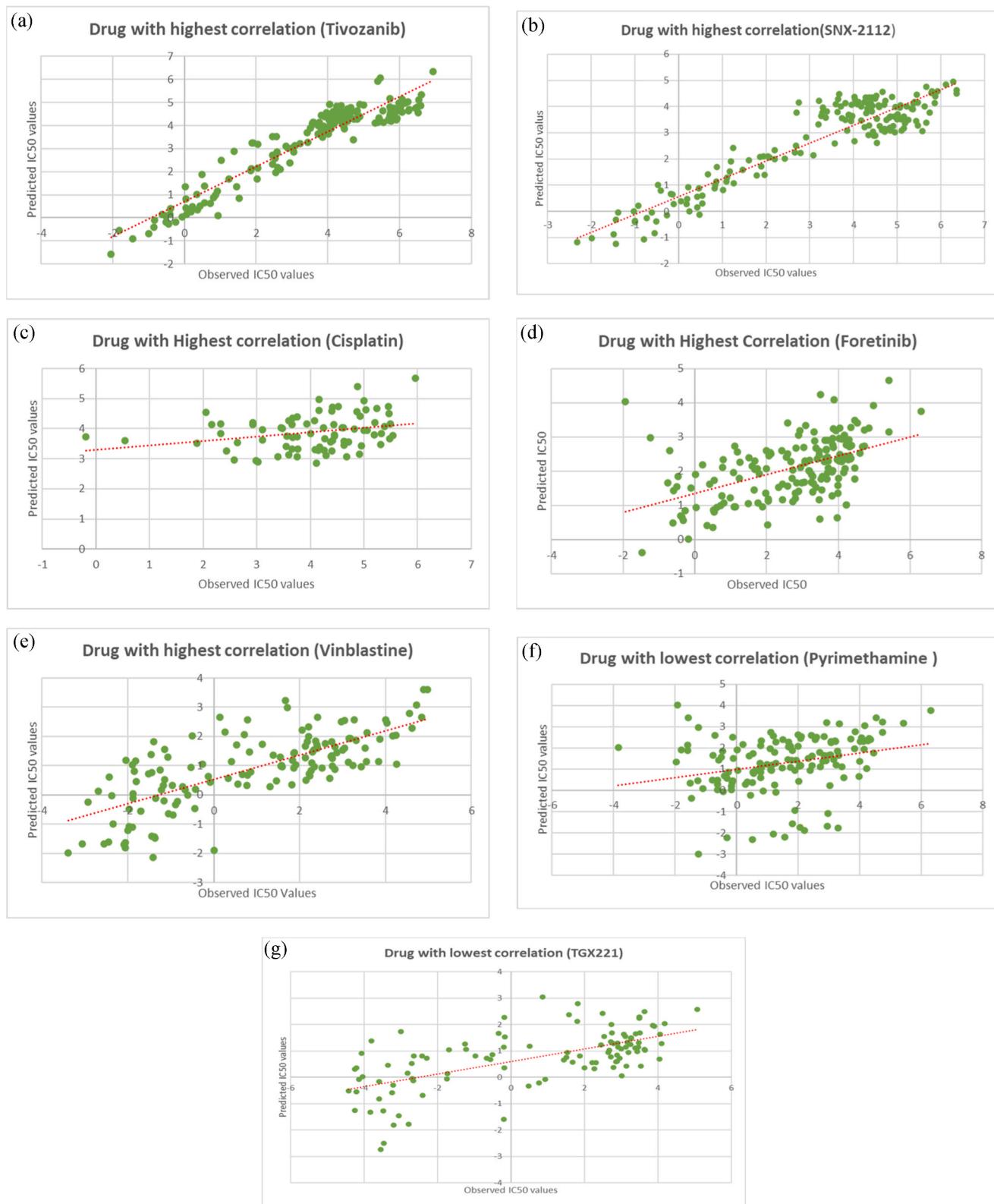
We also examined the drugs that showed the highest performance in different methods. Belinostat showed the most impressive performance in both the DeepCDR and TSGCNN methods. Similarly, Vinblastine outperformed others in DeepTTA. Tivozanib and GFN-2 displayed the highest performance in NIHGCN and TGSA, respectively. Parthenolide emerged as the top performer in the tCNN method. Details of drugs exhibiting best and worst performance in the case of similar methods are depicted in the scatter plots shown in Supplementary File - Figs. 4 and 5, along with their correlation measures. In comparison to the other methods, the current approach demonstrates superior predictive performance, as evidenced by the value of correlation coefficients. Specifically, the highest Pearson's correlation coefficient ( $r_p$ ) of 0.8452 is achieved by the current method for Tivozanib, indicating a strong positive linear relationship between the predicted and observed IC50 values. Additionally, the Spearman's rank correlation coefficient ( $r_s$ ) is 0.8134, reflecting a robust monotonic relationship.

#### 3.4.4. Effect of limited gene data

Several approaches such as DeepTTA (Jiang et al., 2022), TSGCNN (Peng et al., 2023), TGSA (Zhu et al., 2022) etc., utilize the features of the entire gene set (3751 genes (Jin et al., 2023)) in drug response prediction whereas the proposed method uses features of limited genes (697 cancer genes). However, our method achieved competitive results when compared to the methods mentioned below. For example, we compared the results of the proposed method with DeepTTA which utilizes the transcriptomics features of the entire gene dataset. In contrast to the train-test set employed in the DeepTTA method, we allocated 80 % of the known drug-cell line data for training and reserved the remaining 20 % for testing. Notably, certain drugs such as Vinblastine, CMK, and Parthenolide exhibited correlation ( $r_p$ ) of 0.8341, 0.8102 and 0.7833 respectively, between predicted and original IC50 values. Scatter plots of these drugs are presented in Supplementary File, Fig. 6. Subsequently, we compared these findings with the outcomes obtained through our proposed method, which specifically utilizes limited cancer gene data for CDR prediction. Noteworthy drugs like Tivozanib, SNX-2112, Cisplatin, Foretinib, and Vinblastine exhibited significant correlations  $r_p$  of 0.8452, 0.8402, 0.8327, 0.8277 and 0.8231. Corresponding scatter plots of drugs are given in Fig. 6. It is noteworthy that Vinblastine, which exhibited high correlation in the DeepTTA approach, also ranked among the top 5 drugs with high correlations in our method. Furthermore, a widely used anti-cancer medication named Cisplatin, was identified among the top five drugs. This emphasizes the efficacy of our proposed method in generating promising results even when operating with restricted gene data.

#### 3.4.5. Binary classification analysis

Furthermore, the proposed model was extended to a binary classification model to classify the drugs as sensitive or resistant according to the predefined threshold. Since the range of IC50 values is different for individual drug cell-line pairs, to perform binary classification (sensitive or resistant) we used the approach mentioned in Iorio et al. (2016) for binarization. For comparative analysis, binary classification is performed with all the methods under consideration. Metrics used to assess the effectiveness of binary classification models include AUC, AUPR and



**Fig. 6.** a) Scatter Plot of Tivozanib with worst performance by the Proposed method ( $r_p = 0.8452$ ,  $r_s = 0.8134$ ). b) Scatter Plot of SNX-2112 with worst performance by the Proposed method ( $r_p = 0.8402$ ,  $r_s = 0.8022$ ). c) Scatter Plot of Cisplatin with worst performance by the Proposed method ( $r_p = 0.8327$ ,  $r_s = 0.829$ ). d) Scatter Plot of Foretinib with worst performance by the Proposed method ( $r_p = 0.8277$ ,  $r_s = 0.7974$ ). e) Scatter Plot of Vinblastine with worst performance by the Proposed method ( $r_p = 0.8231$ ,  $r_s = 0.7892$ ). f) Scatter Plot of Pyrimethamine with worst performance by the Proposed method ( $r_p = 0.2817$ ,  $r_s = 0.2182$ ). g) Scatter Plot of TGX221 with worst performance by the Proposed method ( $r_p = 0.3023$ ,  $r_s = 0.2844$ ).

ROC Curve. They offer insights into the accuracy with which a model can categorize the two classes (Sensitive or resistant) based on the predictions. These metrics provide a comprehensive evaluation of the model's ability to discriminate between sensitivity and resistance, offering insights into its precision and recall characteristics.

When compared with other similar methods, there are significant improvements in the AUC and AUPR values of the current method. Details of the analysis results are provided in **Table 7**. The proposed method achieved the highest AUC and AUPR scores of 0.7623 and 0.7691 when compared with other similar methods. This indicates that the model effectively distinguishes between positive and negative instances emphasizing its ability to accurately identify positive instances while minimizing false positives.

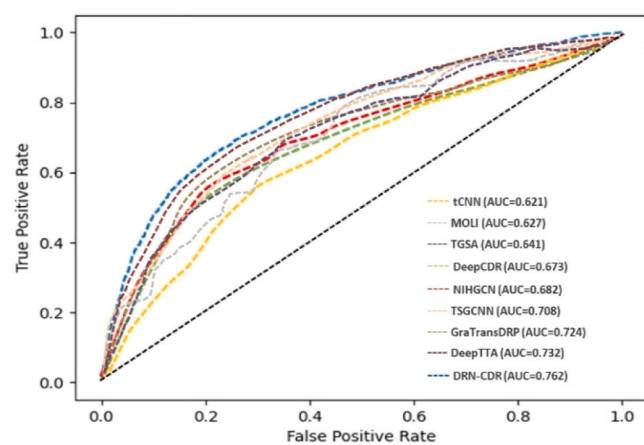
The ROC curve is the graphical representation used to assess the predictive power of the methods. **Fig. 7** shows the ROC curve produced by the proposed method. The farther the curve deviates from the diagonal line, the higher is the model's performance. The ROC curve of the proposed method illustrates its outperformance compared to other models.

### 3.5. Ablation study

In this method, we have integrated multi-omics data and drug features for drug response prediction of drug-cell line pairs. To identify the contribution of individual components to the overall performance of the proposed method, we conduct an ablation study. The impact of individual omics data is analysed by systematically excluding other omics data and drug features. When single omics data was utilized, the  $r_p$  between the observed and predicted response values range between 0.7375 and 0.717. However, the proposed method showed a better  $r_p$  value of 0.7465 when multi-omics and drug features were integrated. Detailed analysis is given in **Table 8**. When comparing the performance of the proposed method among the cases of individual omics data, Transcriptomics (gene expressions) data contributes the most to the performance of the model. From the observations, it can be inferred that the model's performance can be enhanced with the integration of multi-omics data with drug molecular features.

### 3.6. Case studies

To identify whether the model's performance is consistent across different cancer types or if there are any variations in the predictive accuracy, the correlation between predicted and observed IC50 values across each cancer type in TCGA is computed. **Table 8** shows the statistics of predictions corresponding to each cancer type. From **Table 8**, it is evident that the model performs better for TCGA cancer types such as Bladder Urothelial Carcinoma (BLCA), Lung squamous cell carcinoma (LUSC), Liver hepatocellular carcinoma (LIHC) etc., exhibiting the highest correlation values of  $r_p$  above a threshold of 0.9. Therefore, BLCA, LUSC, and LIHC were chosen for the case study to assess the potential of the proposed model in predicting the IC50 values of



**Fig. 7.** ROC curve.

**Table 8**

Ablation study results.

| Components used                             | $r_p$  | $r_s$  | $R^2$  |
|---|--------|--------|--------|
| Genomic<br>(Mutation data)                  | 0.7125 | 0.6577 | 0.6436 |
| Epigenomic<br>(Methylation data)            | 0.7183 | 0.6493 | 0.6691 |
| Transcriptomics<br>(Gene expression data)   | 0.7286 | 0.6804 | 0.6520 |
| Multi-omics without drug molecular features | 0.7588 | 0.7465 | 0.7417 |
| Multi-omics with drug molecular features    | 0.7938 | 0.7804 | 0.7639 |

drug-cell line pairs.

Upon examination of the predictions, it was found that among the specified cancer types, SNX-2112 displayed the lowest median IC50 value compared to other drugs as depicted in **Fig. 8(a)–(c)**. The evidence in the medical literatures such as [Wang et al. \(2015, 2021\)](#), [Xie et al. \(2022\)](#), [Jiang et al. \(2022\)](#) proved that the SNX-2112 is effective for BLCA and other cancers including LUSC. The predictive performance across the cancer types depicted in **Table 8** indicates that drugs such as SNX-2112, Cisplatin, Foretinib, CMK, Vinblastine, BI-2536 and A-4436 showed the lowest median IC50 values in multiple cancer types. Additionally, the medical records corresponding to each cancer type provided in **Table 9** indicate the impact of these drugs in cancer treatment.

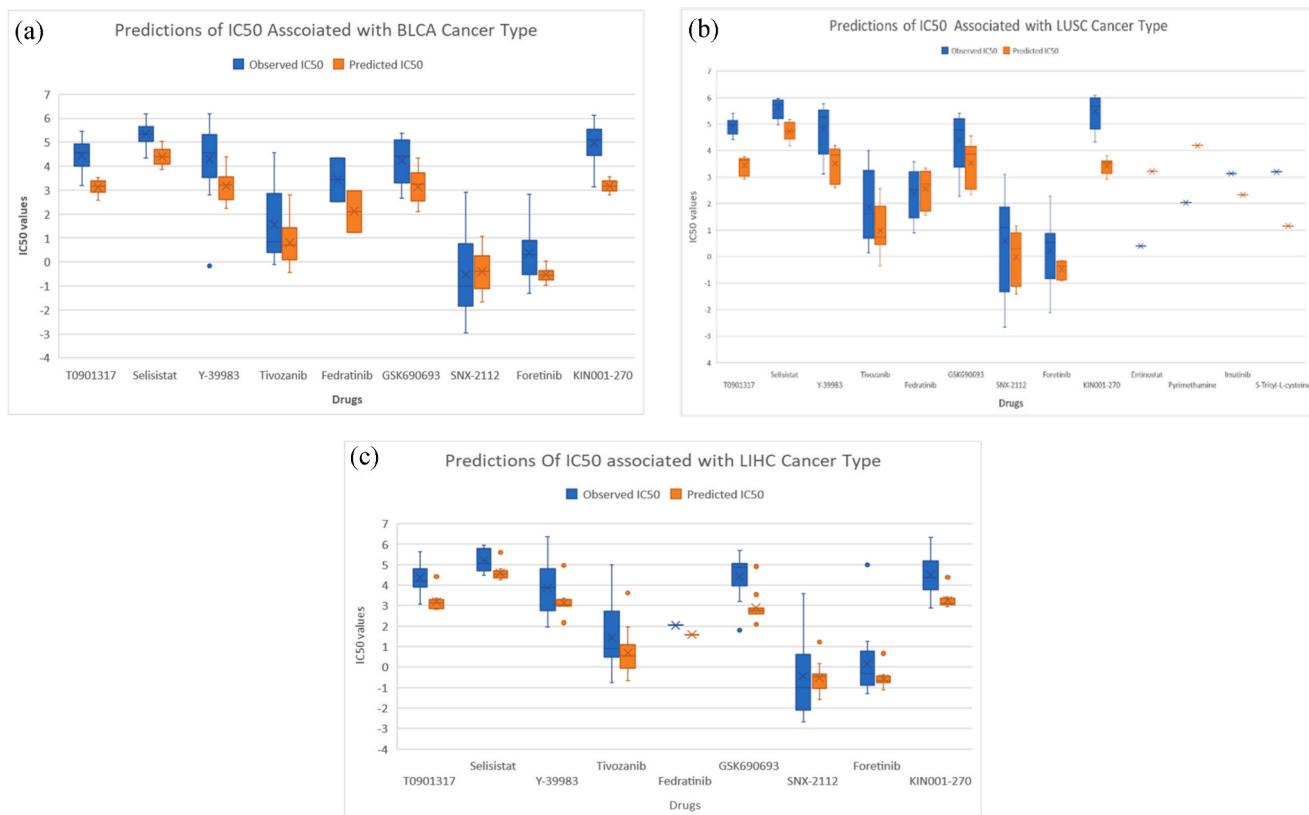
## 4. Discussion

In this work, a Deep ResNet-based cancer drug response prediction model is proposed which integrates drug features (molecular drug structures) and Cell line features (genomics, transcriptomics and epigenomics data). To efficiently capture the input features, UGCN, CNN and fully connected networks were employed that extract feature representation. These features were finally concatenated and fed into a Deep ResNet for IC50 predictions. Since the application of ResNet is relatively underexplored for drug response prediction, we aim to explore its potential in this study. Furthermore, we concentrate specifically on the analysis of Cancer Gene (from COSMIC cancer gene census) datasets, aiming to elucidate their potential influence in predicting cancer drug responses. By narrowing our focus to these datasets, we seek to comprehensively explore the role of cancer-related genetic factors in influencing how cells respond to various treatments. A comparative analysis was carried out to evaluate the predictive power and applicability of the model. The observations from the analysis are summarized below.

- We utilized five random sets for cross-validation, choosing the one that produced the best results to train similar methods for

**Table 7**  
Comparison of performance measures from binary.

| Performance measures                                       | AUC           | AUPR          | Precision     | Accuracy     |
|--|---------------|---------------|---------------|--------------|
| tCNNS ( <a href="#">Liu et al., 2019</a> ) (2019)          | 0.621         | 0.613         | 0.720         | 0.682        |
| MOLI ( <a href="#">Sharifi-Noghabi et al., 2019</a> ) 2019 | 0.627         | 0.615         | 0.736         | 0.709        |
| DeepCDR ( <a href="#">Liu et al., 2020</a> ) (2020)        | 0.673         | 0.656         | 0.744         | 0.736        |
| TGSA ( <a href="#">Zhu et al., 2022</a> ) 2021             | 0.641         | 0.681         | 0.715         | 0.684        |
| NIHGCN ( <a href="#">Peng et al., 2022</a> ) (2022)        | 0.682         | 0.663         | 0.788         | 0.752        |
| DeepTTA ( <a href="#">Jiang et al., 2022</a> ) (2022)      | 0.732         | 0.711         | 0.750         | 0.798        |
| GraTransDRP ( <a href="#">Chu et al., 2022</a> ) (2022)    | 0.724         | 0.697         | 0.737         | 0.783        |
| TSGCNN ( <a href="#">Peng et al., 2023</a> ) (2023)        | 0.708         | 0.675         | 0.713         | 0.791        |
| DRN-CDR  | <b>0.7623</b> | <b>0.7691</b> | <b>0.7704</b> | <b>0.756</b> |



**Fig. 8.** a): Comparison of median IC50 values predictions of drugs for BLCA Cancer Type. b): Comparison of median IC50 values predictions of drugs for LUSC Cancer Type. c): Comparison of median IC50 values of drugs for LIHC Cancer Type.

**Table 9**  
Predictive performance across different cancer types.

| Sl. No. | TCGA label | Cancer types                               | $r_p$       | Drug-Cell line Count | Drug with Lowest median IC50 predictions | Evidence                               |
|---------|------------|--|-------------|----------------------|--|--|
| 1       | BLCA       | Bladder Urothelial Carcinoma               | 0.92751909  | 146                  | SNX-2112                                 | (Wang et al., 2021; Liu et al., 2018a) |
| 2       | LUSC       | Lung squamous cell carcinoma               | 0.916244667 | 78                   | SNX-2112                                 | (Kataoka et al., 2012)                 |
| 3       | LIHC       | Liver hepatocellular carcinoma             | 0.909462312 | 100                  | SNX-2112                                 | (Wang et al., 2014)                    |
| 4       | PAAD       | Pancreatic adenocarcinoma                  | 0.905517011 | 204                  | SNX-2112                                 | (Daunys et al., 2019)                  |
| 5       | THCA       | Thyroid carcinoma                          | 0.903245819 | 75                   | CGP-60474                                | -                                      |
| 6       | HNSC       | Head and Neck Squamous Cell Carcinoma      | 0.828945037 | 148                  | SNX-2112                                 | (Friedman et al., 2013)                |
| 7       | OV         | Ovarian serous cystadenocarcinoma          | 0.783537412 | 216                  | Vinblastine                              | (Du et al., 2018)                      |
| 8       | LUAD       | Lung adenocarcinoma                        | 0.77838681  | 456                  | CGP-60474                                | (Han et al., 2018)                     |
| 9       | BRCA       | Breast invasive carcinoma                  | 0.768025796 | 395                  | Cisplatin                                | (Rodler et al., 2023)                  |
| 10      | MESO       | Mesothelioma                               | 0.740024769 | 39                   | CGP-60474                                | -                                      |
| 11      | KIRC       | Kidney renal clear cell carcinoma          | 0.672367862 | 129                  | CGP-60474                                | -                                      |
| 12      | STAD       | Stomach adenocarcinoma                     | 0.640468462 | 167                  | Foretinib                                | (Grojean et al., 2021)                 |
| 13      | COAD/READ  | Colon adenocarcinoma/Rectum adenocarcinoma | 0.637402607 | 357                  | BI-2536                                  | -                                      |
| 14      | SKCM       | Skin Cutaneous Melanoma                    | 0.621533049 | 330                  | CGP-60474                                | (Yang et al., 2021)                    |
| 15      | GBM        | Glioblastoma Multiforme                    | 0.608091055 | 214                  | CGP-60474                                | -                                      |
| 16      | ALL        | Adult acute lymphoblastic leukemia         | 0.60661333  | 192                  | CMK                                      | -                                      |
| 17      | DLBC       | Diffuse large B cell lymphoma              | 0.582822663 | 270                  | BI-2536                                  | (Vose et al., 2013)                    |
| 18      | LGG        | Brain Lower Grade Glioma                   | 0.532360405 | 73                   | Foretinib                                | -                                      |
| 19      | MM         | Multiple myeloma                           | 0.503760499 | 123                  | A-443654                                 | -                                      |
| 20      | ESCA       | Esophageal carcinoma                       | 0.48645661  | 216                  | Cisplatin                                | (Papadakos et al., 2024)               |
| 21      | LAML       | Acute Myeloid Leukemia                     | 0.47017856  | 225                  | A-443654                                 | (Falà et al., 2008)                    |
| 22      | LCML       | Chronic Myelogenous Leukemia               | 0.447515995 | 99                   | BI-2536                                  | -                                      |
| 23      | SCLC       | Small Cell Lung Cancer                     | 0.425097411 | 367                  | A-443654                                 | (Wildey et al., 2014)                  |
| 24      | NB         | Neuroblastoma                              | 0.41821643  | 108                  | BI-2536                                  | -                                      |

comparison. The Pearson's correlation between the observed and predicted values of IC50 in each validation set is illustrated in [Supplementary File, Fig. 7](#). The proposed method outperforms other

methods across various validation sets. Hence, we choose that set in which the proposed method exhibits the highest correlation for further study.

- The proposed strategy outperformed other competing approaches with the highest Pearson's correlation ( $r_p$ ) and lowest RMSE during the repetitions of the regression exercise. The  $r_p$  was improved to 0.7938, and RMSE was dropped to 0.923. Other similar methods like tCNN, DeepCDR, TGSA, NIHGCN, TSGCNN, GraTransDPR, MOLI and DeepTTA showed  $r_p$  values of 0.65422, 0.61736, 0.68334, 0.68923, 0.72762, 0.76841, 0.79265, 0.76275 and 0.72943, respectively. Additionally, the model yielded the lowest Root Mean Squared Error (RMSE) of 0.9236, in contrast to other methods, where RMSE values ranged from 0.9383 to 1.635.
- The drugs Tivozanib, SNX-2112 and CGP-60474 were found to be the most effective cancer drugs with the lowest median IC<sub>50</sub> value. The other drugs such as PHA-665752, Foretinib etc., were also found to be more sensitive exhibiting the lowest median IC<sub>50</sub> values among other drugs. Medical literatures also corroborated the results.
- The highest value of Pearson's correlation coefficient ( $r_p$ ) achieved for the current method was 0.8452, associated with the drug named Tivozanib and the lowest value was 0.28 corresponding to Pyrimethamine.
- The model was also modified to perform binary classification to categorize the drugs into sensitive or resistant. The proposed method achieved higher AUC and AUPR scores of 0.7623 and 0.7691 when compared to other similar methods. It can also be observed that the ROC curve of the proposed method illustrates better performance than the other models.
- The case studies conducted on BLCA, LUCS and LIHC cancer types consistently reveal that SNX-2112 exhibits the lowest median IC<sub>50</sub> values. Cisplatin, Foretinib, CMK, Vinblastine etc., also showed high sensitivity for other cancer types.

When comparing different methods of drug response prediction, we employed a paired t-test to identify the best-performing approach (Wang et al., 2017). We visualized the Pearson correlation coefficient ( $r_p$ ) between predicted and observed response values of cell lines for each drug using box plots given in Fig. 9. The paired t-test is essential as it determines whether there is a statistically significant difference between the means of the methods being compared. The formula used for paired t-test is given in Eq. (27).

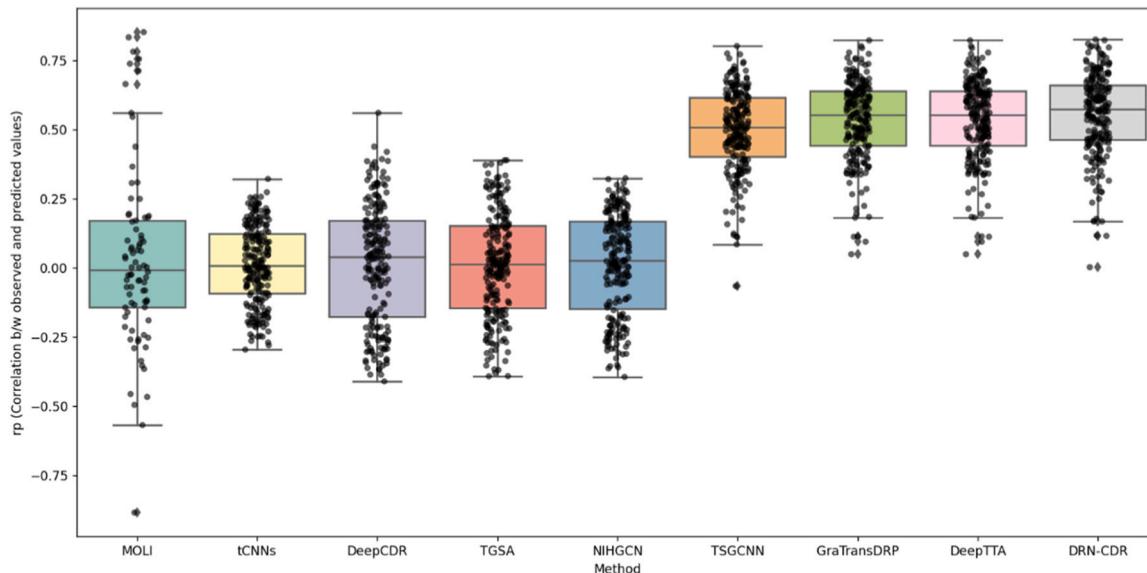
$$t_{value} = \frac{d_{pair'}}{\sigma_{pair}/\sqrt{n}} \quad (27)$$

Where  $d_{pair'}$  is the mean difference between observation of the pair,  $\sigma_{pair}$  is the standard deviation of difference between observations of the pair and n is the number of pairs. The p-value is calculated with degree of freedom as n-1. The p-values between two different methods obtained after the t-test are mentioned in Supplementary File – Table 7. A p-value less than the significance level (i.e., 0.05) indicates a statistically significant difference between the methods. As per the results of paired t-test, DRN-CDR emerged as the best performing method among other compared methods with a comparatively low p-value. Thus, it can be concluded that the DRN-CDR outperforms similar CDR prediction models. However, there exist some limitations which are listed below:

- Exclusive use of Cancer Gene data (COSMIC) contains the risk of overlooking potential targets or failing to detect alterations specific to certain tumor subtypes.
- In the present study, we have not considered information such as metabolomics, proteomics etc., associated with cell lines which may contribute to better CDR prediction. Therefore, further investigation is required to integrate these data for CDR prediction.
- The Pub-Chem database has molecular structures for only 238 drugs, while GDSC offers information on 297 known drugs. We used only the drugs with available molecular structures in our study. As a result, there is a risk of disregarding important drugs.
- The proposed model cannot reliably predict the response for drug combinations. Amending the model to handle drug combinations consuming limited data is another crucial challenge.

## 5. Conclusion and future scope

The focus of this study is to explore the potential of ResNet and exclusive omics data of Cancer genes in CDR prediction. This study proposed a Deep ResNet-based regression model to predict the IC<sub>50</sub> values of drug cell-line pairs using both drug molecular structures and cell-line features. A 1D CNN that handles genomic data and two fully connected networks that handle cell-line-specific transcriptomics and epigenomic features collectively constitute the architecture for cell line feature extraction. A unified graph convolution network (UGCN) is used for drug feature extraction where each drug is represented as a network of constituent atoms. The features of these atoms and the adjacency matrix of the network become the input to the UGCN. The resultant feature representations were concatenated together and supplied to a deep ResNet architecture for predicting IC<sub>50</sub> values of Drug-Cell line



**Fig. 9.** Box plot of different methods with respect to Pearson's correlation between observed and predicted drug responses of different cell lines for each drug.

pairs. Combining drug and cell line features enhances the predictive capability of the model for drug response prediction. The experimental finding provides support for a deeper exploration of the relationships between drug and cellular characteristics and has the potential to influence both research and clinical applications in the field of drug response prediction.

The performance of the proposed model is compared with similar methods. The values of  $r_p$ ,  $r_s$ ,  $R^2$  and RMSE were observed and the proposed model outperformed by showing a greater Pearson's correlation coefficient of 0.7938 between the predicted and the observed IC<sub>50</sub> values. The RMSE of 0.92 in the proposed method is found to be the lowest among the other methods. For classifying the drugs as sensitive or resistant, the proposed model had been modified to a binary classification model. It was observed that the proposed method's ROC curve exhibits superior performance than the other models. The AUC and AUPR measures were 0.7623 and 0.7691 respectively.

It can be concluded that the model is reliable in predicting the IC<sub>50</sub> values of unknown drug-cell line pairs. These improvements suggest that the model has a good ability to estimate drug sensitivity quantitatively. The model demonstrates its versatility as it can be successfully adapted to a binary classification problem. The competitive results achieved when compared with the existing literature strengthen its reliability and potential impact.

The model can further be extended to predict the responses of drug combinations in future. Databases like ComboDB, PubChem, DrugComb etc., can be used to collect drug combination responses. Extend the model by concatenating different drug features or incorporating interaction terms and use transfer learning to fine-tune a pre-trained single-drug model with drug combination data (Chen et al., 2016; Li et al., 2023). Improve predictions by integrating biological knowledge, such as pathway information and synergy scores. This approach leverages existing knowledge and advanced modelling techniques effectively. However, a limitation is that the available drug combination response data from these databases may be limited.

#### CRediT authorship contribution statement

**E R Vimina:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Saranya K R:** Writing – original draft, Software, Investigation, Formal analysis.

#### Author agreement statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

#### 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.

#### References

- Ahmed, Khandakar Tanvir, et al., 2020. Network-based drug sensitivity prediction. *BMC Med. Genom.* 13 (11), 1–10.

- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Garraway, L.A., et al., 2012. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483 (7391), 603–607.
- Botane, Alexandra, Gonçalves, Anthony, Ballester, Pedro J., 2019. Paclitaxel response can be predicted with interpretable multi-variate classifiers exploiting DNA-methylation and miRNA data. *Front. Genet.* 10, 1041.
- Broussy, Sylvain, Laaroussi, Hanna, Vidal, Michel, 2020. Biochemical mechanism and biological effects of the inhibition of silent information regulator 1 (SIRT1) by EX-527 (SEN0014196 or selisistat). *J. Enzym. Inhib. Med. Chem.* 35 (1), 1124–1136.
- Chang, Yoosup, et al., 2018. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci. Rep.* 8 (1), 8857.
- Chen, Siqi, et al., 2023. DNN-PNN: a parallel deep neural network model to improve anticancer drug sensitivity. *Methods* 209, 1–9.
- Chen, Xing, et al., 2016. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. *PLoS Comput. Biol.* 12 (7), e1004975.
- Chiu, Yu-Chiao, et al., 2019. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genom.* 12 (1), 143–155.
- Chu, Thang, et al., 2022. Graph transformer for drug response prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 20 (2), 1065–1072.
- Daunys, Simona, Matulis, Daumantas, Petrikaité, Vilma, 2019. Synergistic activity of Hsp90 inhibitors and anticancer agents in pancreatic cancer cell cultures. *Sci. Rep.* 9 (1), 16177.
- Dhyani, Praveen, et al., 2022. Anticancer potential of alkaloids: a key emphasis to colchicine, vinblastine, vincristine, vindesine, vinorelbine and vincamine. *Cancer Cell Int.* 22 (1), 206.
- Ding, Z., Zu, S., Gu, J., 2016. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* 32 (19), 2891–2895.
- Du, Guan-Hua, et al., 2018. Vinblastine and vincristine. *Nat. Small Mol. Drugs Plants* 551–557.
- Emdadi, Akram, Eslahchi, Changiz, 2020. Dsplmf: a method for cancer drug sensitivity prediction using a novel regularization approach in logistic matrix factorization. *Front. Genet.* 11, 509293.
- Eskens, Ferry A.L.M., et al., 2011. Biologic and clinical activity of tivozanib (AV-951, KRN-951), a selective inhibitor of VEGF receptor-1, -2, and -3 tyrosine kinases, in a 4-week-on, 2-week-off schedule in patients with advanced solid tumors. *Clin. Cancer Res.* 17 (22), 7156–7163.
- Falà, Federica, et al., 2008. Proapoptotic activity and chemosensitizing effect of the novel Akt inhibitor (2S)-1-(1H-Indol-3-yl)-3-[5-(3-methyl-2H-indazol-5-yl) pyridin-3-yl] oxypyran-2-amine (A443654) in T-cell acute lymphoblastic leukemia. *Mol. Pharm.* 74 (3), 884–895.
- Friedman, Jay A., et al., 2013. HSP90 inhibitor SNX5422/2112 targets the dysregulated signal and transcription factor network and malignant phenotype of head and neck squamous cell carcinoma. *Transl. Oncol.* 6 (4), 429–INS.
- Gao, Galen F., et al., 2019. Before and after: comparison of legacy and harmonized TCGA genomic data commons' data. *Cell Syst.* 9 (1), 24–34.
- Grojean, Meghan, et al., 2021. Targeted dual inhibition of c-Met/VEGFR2 signalling by foretinib improves antitumour effects of nanoparticle paclitaxel in gastric cancer models. *J. Cell. Mol. Med.* 25 (11), 4950–4961.
- Guan, Na-Na, et al., 2019. Anticancer drug response prediction in cell lines using weighted graph regularized matrix factorization. *Mol. Ther.-Nucleic Acids* 17, 164–174.
- Güvenç Paltun, Betül, Mamitsuka, Hiroshi, Kaski, Samuel, 2021. Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches. *Brief. Bioinform.* 22 (1), 346–359.
- Han, E.K.H., et al., 2007. Akt inhibitor A-443654 induces rapid Akt Ser-473 phosphorylation independent of mTORC1 inhibition. *Oncogene* 26 (38), 5655–5661.
- Han, Hyun-Wook, et al., 2018. LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxicemic agent. *Sci. Rep.* 8 (1), 14969.
- Haque, Ashamul, et al., 2018. Next generation antineoplastic agents: a review on structurally modified vinblastine (VBL) analogues. *Curr. Med. Chem.* 25 (14), 1650–1662.
- He, Kaiming, et al., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Hermida, Leandro C., Gertz, E.Michael, Rupp, Eytan, 2022. Predicting cancer prognosis and drug response from the tumor microbiome. *Nat. Commun.* 13 (1), 2896.
- Huang, Shujun, Hu, Pingzhao, Lakowski, Ted M., 2021. Predicting breast cancer drug response using a multiple-layer cell line drug response network model. *BMC Cancer* 21 (1), 648.
- Iorio, F., et al., 2016. A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754.
- Jiang, Likun, et al., 2022. DeepTTA: a transformer-based model for predicting cancer drug response. *Brief. Bioinform.* 23 (3), bbac100.
- Jin, Han, et al., 2023. Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nat. Commun.* 14 (1), 5417.
- Kataoka, Yu, et al., 2012. Foretinib (GSK1363089), a multi-kinase inhibitor of MET and VEGFRs, inhibits growth of gastric cancer cell lines by blocking inter-receptor tyrosine kinase networks. *Investig. New Drugs* 30, 1352–1360.
- Koras, Krzysztof, et al., 2020. Feature selection strategies for drug sensitivity prediction. *Sci. Rep.* 10 (1), 9377.
- Lee, Dong Shin, et al., 2024. Engineered 3D tumor microenvironment recapitulating stiffness of lung tissue to explore drug resistance of lung carcinoma. *J. Ind. Eng. Chem.* 132, 360–368.
- Li, Tian-Hao, et al., 2023. SNRMPACDC: computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction. *Brief. Bioinform.* 24 (1), bbac503.

- Li, Yuanyuan, et al., 2021. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genom.* 22, 1–18.
- Liu, Yuting, et al., 2015. Combination of SNX-2112 with 5-FU exhibits antagonistic effect in esophageal cancer cells. *Int. J. Oncol.* 46 (1), 299–307.
- Liu, Hui, et al., 2018b. Anti-cancer drug response prediction using neighbor-based collaborative filtering with global effect removal. *Mol. Ther.-Nucleic Acids* 13, 303–311.
- Liu, Pengfei, et al., 2019. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinform.* 20 (1), 1–14.
- Liu, Qiao, et al., 2020. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 36 (Suppl. 2), Si911–Si918.
- Liu, Xuan, et al., 2018a. Targeted delivery of SNX-2112 by polysaccharide-modified graphene oxide nanocomposites for treatment of lung cancer. *Carbohydr. Polym.* 185, 85–95.
- Liu, Xuan, et al., 2022. GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction. *Brief. Bioinform.* 23 (1), bbab457.
- Liu, Xuan, Zhang, Wen, 2023. A subcomponent-guided deep learning method for interpretable cancer drug response prediction. *PLoS Comput. Biol.* 19 (8), e1011382.
- Meybodi, Fatemeh Yassaei, Eslahchi, Changiz, 2021. Predicting anti-cancer drug response by finding optimal subset of drugs. *Bioinformatics* 37 (23), 4509–4516.
- Moughari, Fatemeh Ahmadi, Eslahchi, Changiz, 2020. ADRML: anticancer drug response prediction using manifold learning. *Sci. Rep.* 10 (1), 14245.
- Nazari, Somayeh, et al., 2024. Foretinib, a c-MET receptor tyrosine kinase inhibitor, tackles multidrug resistance in cancer cells by inhibiting ABCB1 and ABCG2 transporters. *Toxicol. Appl. Pharm.*, 116866.
- Papadakos, Stavros P., et al., 2024. Metformin in esophageal carcinoma: exploring molecular mechanisms and therapeutic insights. *Int. J. Mol. Sci.* 25 (5), 2978.
- Parca, Luca, et al., 2019. Modeling cancer drug response through drug-specific informative genes. *Sci. Rep.* 9 (1), 15222.
- Peng, Wei, et al., 2022. Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions. *Bioinformatics* 38 (19), 4546–4553.
- Peng, Wei, et al., 2023. Improving drug response prediction based on two-space graph convolution. *Comput. Biol. Med.* 158, 106859.
- Peng, Wei, Chen, Tielin, Dai, Wei, 2021. Predicting drug response based on multi-omics fusion and graph convolution. *IEEE J. Biomed. Health Inform.* 26 (3), 1384–1393.
- Rampásek, Ladislav, et al., 2019. Dr. VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics* 35 (19), 3743–3751.
- Ramsundar, B., Eastman, P., Walters, P., Pande, V., 2019. Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More. O'Reilly Media.
- Rodler, Eve, et al., 2023. Cisplatin with veliparib or placebo in metastatic triple-negative breast cancer and BRCA mutation-associated breast cancer (S1416): a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Oncol.* 24 (2), 162–174.
- Romani, Andrea M.P., 2022. Cisplatin in cancer treatment. *Biochem. Pharm.* 206, 115323.
- Saginalieva, Asel, et al., 2023. Hybrid quantum neural network for drug response prediction. *Cancers* 15 (10), 2705.
- Santoni, Matteo, et al., 2018. Tivozanib for the treatment of renal cell carcinoma. *Expert Opin. Pharmacother.* 19 (9), 1021–1025.
- Schöffski, Patrick, et al., 2010. Multicentric parallel phase II trial of the polo-like kinase 1 inhibitor BI 2536 in patients with advanced head and neck cancer, breast cancer, ovarian cancer, soft tissue sarcoma and melanoma. The first protocol of the European Organization for Research and Treatment of Cancer (EORTC) Network Of Core Institutes (NOCI). *Eur. J. Cancer* 46 (12), 2206–2215.
- 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.
- Sohn, Sung-Hwa, et al., 2020. Foretinib inhibits cancer stemness and gastric cancer cell proliferation by decreasing CD44 and c-MET signaling. *OncoTargets Ther.* 1027–1035.
- Suphavilai, Chayaporn, Bertrand, Denis, Nagarajan, Niranjan, 2018. Predicting cancer drug response using a recommender system. *Bioinformatics* 34 (22), 3907–3914.
- Taguchi, Y.H., 2019. Drug candidate identification based on gene expression of treated cells using tensor decomposition-based unsupervised feature extraction for large-scale data. *BMC Bioinform.* 19, 27–42.
- Tchounwou, Paul B., et al., 2021. Advances in our understanding of the molecular mechanisms of action of cisplatin in cancer therapy. *J. Exp. Pharm.* 303–328.
- Vose, Julie M., et al., 2013. "The Plk1 inhibitor BI 2536 in patients with refractory or relapsed non-Hodgkin lymphoma: a phase I, open-label, single dose-escalation study. *Leuk. Lymphoma* 54 (4), 708–713.
- Wang, Xiao, et al., 2015. Comparative effects of SNX-7081 and SNX-2112 on cell cycle, apoptosis and Hsp90 client proteins in human cancer cells. *Oncol. Rep.* 33 (1), 230–238.
- Wang, Lin, et al., 2017. Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* 17, 1–12.
- Wang, Shiming, Li, Jie, 2020. Modular within and between score for drug response prediction in cancer cell lines. *Mol. Omics* 16 (1), 31–38.
- Wang, Xiao, et al., 2014. The Hsp90 inhibitor SNX-2112 induces apoptosis of human hepatocellular carcinoma cells: the role of ER stress. *Biochem. Biophys. Res. Commun.* 446 (1), 160–166.
- Wang, Zhi, et al., 2021. Identification of a tumor microenvironment-related seven-gene signature for predicting prognosis in bladder cancer. *BMC Cancer* 21 (1), 1–18.
- Wildey, Gary, et al., 2014. Pharmacogenomic approach to identify drug sensitivity in small-cell lung cancer. *PLoS One* 9 (9), e106784.
- Xia, Fangfang, et al., 2022. A cross-study analysis of drug response prediction in cancer cell lines. *Brief. Bioinform.* 23 (1), bbab356.
- Xie, Ruiyan, et al., 2022. The relationship of pyroptosis-related genes, patient outcomes, and tumor-infiltrating cells in bladder urothelial carcinoma (BLCA). *Front. Pharm.* 13, 930951.
- Yadav, Anil Kumar, et al., 2024. PHA-665752's antigrowth and proapoptotic effects on hsc-3 human oral cancer cells. *Int. J. Mol. Sci.* 25 (5), 2871.
- Yang, Ronghua, et al., 2021. The identification of the metabolism subtypes of skin cutaneous melanoma associated with the tumor microenvironment and the immunotherapy. *Front. Cell Dev. Biol.* 9, 707677.
- Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Garnett, M.J., et al., 2012. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41 (D1), D955–D961.
- Yuan, Yixiao, et al., 2021. SNX20AR/MiRNA-301a-3p/SNX20 axis associated with cell proliferation and immune infiltration in lung adenocarcinoma. *Front. Mol. Biosci.* 8, 744363.
- Zhang, Lin, et al., 2018. A hybrid interpolation weighted collaborative filtering method for anti-cancer drug response prediction. *Front. Pharm.* 9, 1017.
- Zhidkova, E.M., et al., 2022. In vitro screening of effectiveness and antiproliferative effects of potential DDIT4 inhibitors for breast cancer cell lines. *Sib. J. Oncol.* 21 (3), 50–60.
- Zhu, Yiheng, et al., 2022. TGSA: protein–protein association-based twin graph neural networks for drug response prediction with similarity augmentation. *Bioinformatics* 38 (2), 461–468.

**Saranya K. R.** Received Bachelor's Degree in Computer Science and Master's Degree in Computer Science from University of Calicut, Kerala, India. Currently doing her Doctoral Research in the field of Computational Biology under the Department of Computer Science and IT, School of Computing, Amrita Vishwa Vidyapeetham-Kochi Campus. Research interests includes Deep Learning, Computational Biology and Bioinformatics.

**Vimina E. R.** received B. Tech degree in Electrical and Electronics Engineering from Mahatma Gandhi University in 1999, M.E. degree in Computer Science and Engineering from Bharathiar University in 2000 and the Ph.D. degree in Computer Science from Cochin University of Science and Technology in 2018. She is currently working as Assistant Professor in the Department of Computer Science and IT, School of Computing, Amrita Vishwa Vidyapeetham, Kochi Campus, India. Her research interests include Computational Biology, Computer Vision and Deep-learning.