

## RESEARCH ARTICLE

## ASGCL: Adaptive Sparse Mapping-based graph contrastive learning network for cancer drug response prediction

Yunyun Dong<sup>1,2</sup>, Yuanrong Zhang<sup>1</sup>, Yuhua Qian<sup>2,3\*</sup>, Yiming Zhao<sup>1</sup>, Ziting Yang<sup>1</sup>, Xiufang Feng<sup>1</sup>**1** School of Software, Taiyuan University of Technology, Taiyuan, China, **2** Institute of Big Data Science and Industry, Shanxi University, Taiyuan, China, **3** School of Computer and Information Technology, Shanxi University, Taiyuan, China\* [jinchengqyh@126.com](mailto:jinchengqyh@126.com)**OPEN ACCESS****Citation:** Dong Y, Zhang Y, Qian Y, Zhao Y, Yang Z, Feng X (2025) ASGCL: Adaptive Sparse Mapping-based graph contrastive learning network for cancer drug response prediction. *PLoS Comput Biol* 21(1): e1012748. <https://doi.org/10.1371/journal.pcbi.1012748>**Editor:** Özlem Tastan Bishop, Rhodes University, SOUTH AFRICA**Received:** March 21, 2024**Accepted:** December 23, 2024**Published:** January 30, 2025**Copyright:** © 2025 Dong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** Supporting source code is available online at <https://github.com/Zymonody77/ASGCL>.**Funding:** This work was supported by the National Natural Science Foundation of China (No. 62306206, funded by Y. D.), the Shanxi Province Key Research and Development Program (No. 202102020101007, funded by X. F.), the Shanxi Province Applied Basic Research Program (No. 202203021212207, funded by Y. D.) and the Taiyuan University of Technology Youth Fund (No.**Abstract**

Personalized cancer drug treatment is emerging as a frontier issue in modern medical research. Considering the genomic differences among cancer patients, determining the most effective drug treatment plan is a complex and crucial task. In response to these challenges, this study introduces the Adaptive Sparse Graph Contrastive Learning Network (ASGCL), an innovative approach to unraveling latent interactions in the complex context of cancer cell lines and drugs. The core of ASGCL is the GraphMorpher module, an innovative component that enhances the input graph structure via strategic node attribute masking and topological pruning. By contrasting the augmented graph with the original input, the model delineates distinct positive and negative sample sets at both node and graph levels. This dual-level contrastive approach significantly amplifies the model's discriminatory prowess in identifying nuanced drug responses. Leveraging a synergistic combination of supervised and contrastive loss, ASGCL accomplishes end-to-end learning of feature representations, substantially outperforming existing methodologies. Comprehensive ablation studies underscore the efficacy of each component, corroborating the model's robustness. Experimental evaluations further illuminate ASGCL's proficiency in predicting drug responses, offering a potent tool for guiding clinical decision-making in cancer therapy.

**Author summary**

Cancer is a globally challenging disease characterized by significant heterogeneity, predicting anti-cancer drug sensitivity plays critical roles in drug development and developing personalized drug treatment plans for cancer. Some computational methods have been proposed to address interactions between multiple cell lines and drugs, and provide promising results. However, the existing methods fail to fully consider the potential complex non-linear relationships and struggle to capture long-range dependencies in the underlying graph. To overcome above questions, a novel computational method named ASGCL model is proposed to amalgamate the characteristics and documented reactions of cancer cell lines and drugs.

2022QN041, funded by Y. D.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

ASGCL constructs graph augmentation module GraphMorpher, and generates enhanced graphs, preserving key structures and attributes while sparsifying secondary features and topologies in each iteration. The experimental results show that ASGCL is able to effectively integrate data of cancer cell lines and drugs with existing response activities.

## Introduction

Cancer's global challenge lies in its vast heterogeneity. Even within the same cancer type, drug responses vary due to genetic differences [1,2,3], emphasizing the urgency of personalized cancer treatments, a core aim in precision medicine. In drug development, predicting anti-cancer drug sensitivity has become a central research focus, involving genomics, transcriptomics, proteomics, and related disciplines [4,5], [6]. Traditional prediction methods, based on statistical matrix decomposition [7], rely on data's statistical features, overlooking uncertainties and incompleteness. This limitation makes it challenging to address interactions between multiple cell lines and drugs, leading to inaccuracies and misjudgments in predictions.

With the development of artificial intelligence, machine learning and deep learning methods for predicting drug resistance are emerging. For instance, classic machine learning algorithms like logistic regression [8], support vector machines [9], and random forests [10] have shown significant effectiveness in drug resistance prediction. These methods are primarily based on single-omics data such as genetic mutations or gene expression. However, integrating multi-omics data allows for the extraction of more advanced correlations, thereby enhancing their predictive capabilities. For instance, NetLapRLS [11] utilizes the Laplacian regularized least squares method, combining known drug-target interactions, chemical structures, and genomic sequence data for drug resistance prediction.

With the advancement of genomics, the application of novel genomic data [4,12,13] in predicting drug resistance [14,15] is increasingly prevalent. For instance, the method proposed by Su et al. [16] transforms cell line gene expression and copy number variations into high-dimensional feature vectors to predict drug responses in corresponding cell lines. However, most current machine learning algorithms model the correlation between cell line features and drug features linearly, failing to fully consider the potential complex non-linear relationships.

Another approach to predicting drug resistance focuses on using deep learning models to extract features of cell lines and drugs and input them into classifiers to predict the cells' responses to drugs [15,16,17,18,19,20,21]. Various deep neural networks, like multilayer perceptrons [22,23], convolutional neural networks [21,22,23,24], and recurrent neural networks [25], have been used to learn potential representations of cancer cell lines and drugs. Hossein et al. [15] encode gene expression, somatic mutations, and copy number variations of cell lines, creating feature representations of these lines. However, most feature-based drug resistance prediction methods overlook the associations between cell lines or drugs. To address this, Zhang et al. [19] established a dual similarity network based on cancer cell line gene expression and drug chemical structures for predicting drug resistance. Turki et al. [26] introduced a link-filtering algorithm based on the cancer cell line network, combined with a linear regression model for the prediction task.

Recent research in bioinformatics shows that graph structures excel in predicting drug resistance, current methods primarily use graph structures for modeling nonlinear structural features of cell lines and molecular structures of drugs. For instance, DeepCDR [21] is a hybrid prediction model based on GCN, integrating data expression. HNMDRP [27] constructs a heterogeneous network of cell line-drug targets and employs a flow-based prediction algorithm.

In general, these graph-based methods extract cell line features from gene expression, somatic mutations, and copy number variations, and combine them with drug FingerPrints and SMILES molecular structures to create molecular graphs for extracting drug features [20]. However, due to the increasing layers in graph neural networks lead to over-smoothing issues, making the final feature representations less distinct and affecting prediction accuracy.

Given the limited availability of detailed annotated data on cancer cell, self-supervised learning approach focuses on automatically extracting features from data. Contrastive learning [28,29,30] aims to enhance sample processing by dividing them into positive and negative sets, striving to maximize the similarity within positive sets and minimize it within negative sets. [30,31,32]. For example, GraphCDR [33] creates contrastive learning tasks for cancer cell line drug resistance reactions, thereby enhancing the model's generalization capability and learning richer feature representations.

In our study, we proposed an Adaptive Sparse Mapping based Graph Contrastive Learning Network (ASGCL) for predicting cancer drug response. Extracting the original features of cell lines and drugs through nonlinear subspaces, mapping them into heterogeneous graphs through cell line drug interactions, and then proposing the GraphMorpher module to perform adaptive graph sparsity on heterogeneous graphs. Finally, a contrastive learning task is designed on multiple graph structures to enhance feature discrimination, and a graph encoder is used to learn potential features of cell lines and drugs from sensitive graphs for supervised prediction tasks. The contributions of this paper are as follows:

- Within the graph contrastive learning framework, the ASGCL model amalgamates the characteristics and documented reactions of cancer cell lines and drugs. By integrating multi-omics information, the predictive performance of cancer drug responses has been significantly improved.
- This study innovatively developed the adaptive graph augmentation module GraphMorpher, which performs masking operations on nodes and pruning operations on links. In each iteration, the module adaptively generates enhanced graphs, preserving key structures and attributes while sparsifying secondary features and topologies.
- Additionally, by implementing contrastive learning tasks, this study greatly improved the model's generalization capability, defining positive and negative samples at the node and graph levels and setting multi-level learning objectives, further enhancing the model's discrimination ability.

## Related work

### Similarity matrices

Numerous statistical-based methods have been proposed for predicting drug resistance, including techniques such as matrix decomposition. For instance, SRMF [7] is a matrix decomposition algorithm that analyzes cell line and drug characteristics by decomposing known cell line-drug relation matrices and predicts drug responses. DTINet [34] learns low-dimensional representations of drugs and proteins from multiple networks and employs inductive matrix completion techniques to predict drug-target interactions. Zhang et al. [27] constructed a heterogeneous network model by calculating the Pearson correlation coefficients between cell line genomic profiles, drug chemical structures, and target genes. However, these methods primarily focus on identifying drug-cell line interactions through linear combinations of latent features. In reality, interactions between drugs and individual cell lines may exhibit significant non-linearity.

Matrix decomposition-based methods [35,36] typically involve multiplying decomposition factors to reconstruct known drug-response relationships, that these factors are constrained by side information from cancer cell lines and drugs [37,38], potentially impacting the task of predicting drug resistance.

## Graph Neural Networks (GNN)

With the capability to learn complex functions and high-dimensional representations continues to improve, particularly, models based on GNN [39,40] are notably significant in exploring relationships between drugs and cell lines due to their effective use of graph topology to aggregate node features.

For instance, GraphDRP [20] obtains the embedded representation of drugs through GCN, while using binary vectors of genomic mutations to acquire the embedded representation of cell lines. NIHGCN [41] constructs a cell line-drug heterogeneous network, using cell line gene expression and drug fingerprints as node features. Wei et al. [42] proposed an end-to-end algorithm named MOFGCN that based on multi-omics integration and GCN to predict drug responses in cell lines.

The above research methods based on graph structures have achieved many academic results. Although most methods consider the relationships between drugs and cell lines when constructing model structures, many methods for modeling cell lines and drug structures do not take into account the global features of graph structures, blindly using multi-layer GCNs, gradually covering the features with the features of message transmission mechanisms, resulting in the convergence of features between nodes with message transmission.

## Contrastive learning

Although existing methods have achieved significant results, most models rely on label-supervised training, which is limited by the size of datasets and the scarcity of task-specific labels in graph-structured datasets.

In response to this challenge, some previous research has introduced contrastive learning in drug resistance tasks. For instance, Yao et al. [43] developed the SHGCL-DTI framework by incorporating a contrastive learning module in semi-supervised prediction tasks. Graph contrastive learning provides a new framework for learning feature representations in graph structures where labeled data is limited or absent. For example, Liu et al. [44] selected k-nearest neighbor nodes as positive samples for each node and selected negative samples from the remaining nodes in the graph.

Furthermore, existing studies [33,34,43,44,45,46,47,48] have not yet fully explored the rich semantic information and complex interactions between data in heterogeneous graphs. Wei et al. [49] proposed a contrastive learning model, which integrating contrastive learning strategies into graph collaborative filtering. Gao et al. [50] introduced a graph joint contrastive learning model, using network topology for learning embeddings. Most of the above studies focus on generating similar positive samples within graph structures to learn similar features, but they often overlook the importance of negative samples.

## Results

### Dataset

We conducted experimental tests of our model on two datasets: GDSC and CCLE. This study utilized the GDSC dataset to determine the sensitivity of cell lines to drugs, defining cell line-drug pairs as sensitive if their IC50 values exceeded the threshold specified in the threshold

table. The GDSC database has 990 cancer cell lines and 265 drugs, with 20,851 sensitive samples and 156,512 resistant samples. For the CCLE dataset, the  $\log_{10}(\text{IC}_{50})$  values of all cell lines were normalized to a zero mean and unit variance. If the Z-score normalized  $\log_{10}(\text{IC}_{50})$  value is less than -0.8, the cell line is defined as sensitive to the drug. Otherwise, the cell line is considered to be resistant to the drug. Ultimately, the CCLE dataset comprised 1,696 sensitive samples and 8,768 resistant samples. We extracted gene expression, copy number variations, and somatic mutation histological features of the respective cell lines from the GDSC. Additionally, we removed from PubChem cell lines lacking any histological data and drugs with identical compound IDs (CIDs). After processing, we obtained the gene expression, copy number variations, and somatic mutation histological features of 962 cell lines, along with their responses to 228 drugs.

## Experimental details

We utilize the GPU of RTX Nvidia 4090, conducting training and testing based on the PyTorch framework. The dataset is divided into training and test sets using five-fold cross-validation, and the model parameters are updated using the Adam optimizer. In the experiments, the learning rate is set to  $5e-4$ ; the number of hidden layers is configured as (962,224), with the hidden layers in the graph encoder set to (224,24). The settings of other parameters will be substantiated in subsequent experiments.

In our model, a nonlinear subspace encoder is constructed, and a selection is made among Graph Neural Network (GNN), Graph Convolutional Network (GCN) and Graph Attention Network (GAT). Ultimately, we chose GCN ( $k = 1$ ) as the nonlinear subspace encoder. The detailed comparison process is provided in the ablation study section.

## Evaluation criteria

This paper evaluates the performance of the model using the Area Under the Receiver Operating Characteristic curve (AUC), Average Precision (AP), Accuracy (ACC), F1 Score, and Matthews Correlation Coefficient (MCC). Mathematically, these metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{ACC} = \frac{TP}{TP + TN} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

where true Positives (TP), True Negatives (TN), False Negatives (FN), and False Positives (FP) are the four categories of the confusion matrix, calculated by comparing predicted values with actual response values. The F1 Score is defined as the harmonic mean of precision and recall. The MCC produces a high score only when the prediction yields good results across all four categories of the confusion matrix (TP, TN, FN, and FP).

## Loss function

In this study, contrastive loss and supervised loss are combined in an end-to-end manner to learn feature representations. The extent of loss for supervised tasks is determined using the cross-entropy loss function, which can be represented as:

$$Loss_{sup} = -\frac{1}{|S|} \sum (P_{lab} \log \tilde{P} + (1 - P_{lab}) \log \tilde{P}) \quad (6)$$

where  $S$  represents the training set of the same batch,  $P_{lab}$  the actual label values between nodes, and  $\tilde{P}$  the predicted values between nodes.

To simultaneously accomplish the task of drug resistance prediction and contrastive learning, the losses from contrastive learning and supervised learning are combined, optimizing the following loss function:

$$Loss = \alpha \cdot Loss_{sup} + \beta Loss_{nod} + \gamma Loss_{gra} \quad (7)$$

where  $\alpha, \beta$  and  $\gamma$  are hyperparameters representing the contributions of different loss functions, with  $\alpha + \beta + \gamma = 1$ . In our study, we set the hyperparameter  $\alpha$  to 0.4 and the hyperparameters  $\beta$  and  $\gamma$  to 0.4 and 0.2, respectively, maintaining a 2:1 ratio.

## Experimental results

To verify the effectiveness of our proposed ASGCL model, multi-method comparative experiments were conducted in this section, along with quantitative and qualitative assessments of these methods. The methods used in the comparative experiments include DeepCDR, DeepDSC, MOFGCN, GraphCDR, and TSGCNN. All comparison methods are based on the same datasets and evaluation metrics, and the default or optimal parameter settings for each model were adopted. All benchmark models utilized the same input features and preprocessing strategies as ASGCL. Stratified 5-fold cross-validation was used to assess model performance. Specifically, the samples in the dataset were divided into five equal parts, each part taking turns as the test set, with the remainder serving as the training set. In each fold, the performance of different models was evaluated using the average of various metrics.

The experimental results, as shown in [Table 1](#), indicate that the ASGCL model achieved the highest performance on both the GDSC and CCLE datasets compared to baseline models. Compared to other models, ASGCL showed superior performance with the highest AUC (96.23%), AP (97.73%), ACC (90.32%), F1 Score (89.66%), and MCC (81.33%), indicating a good match between IC50 values and predicted values. The main reason is that the model proposed in this paper obtained reliable feature representations of cell lines and drugs through contrastive learning on graph structures. In other models, the feature representations were randomly initialized, neglecting the rich biomedical information related to cell lines and drugs. These results demonstrate that the ASGCL model is more effective in extracting additional drug potential data, and the learned embeddings are more representative.

## Classification task results

To further assess the performance of prediction models on the GDSC and CCLE datasets, this study conducted an independent validation, testing the performance of the proposed ASGCL model and comparison models in drug response classification tasks, using ROC curves as the evaluation metric. As hyperparameter tuning and model training are independent of the separate test set, the separate test set can better measure the ASGCL model's generalization ability on unknown data. As shown in [Fig 1](#), [Fig 1A](#) depicts the ROC curve of the ASGCL model on

Table 1. Performance comparison of different algorithms based on five indicators.

Dataset	Algorithm	AUC	AP	ACC	F1_Score	MCC
GDSC	DeepCDR	0.8009±1×10 <sup>-2</sup>	0.8091±1×10 <sup>-2</sup>	0.7748±1×10 <sup>-2</sup>	0.8063±1×10 <sup>-3</sup>	0.5812±1×10 <sup>-2</sup>
	DeepDSC	0.8174±1×10 <sup>-2</sup>	0.8327±1×10 <sup>-3</sup>	0.7901±1×10 <sup>-2</sup>	0.8154±1×10 <sup>-3</sup>	0.6073±1×10 <sup>-2</sup>
	MOFGCN	0.8785±1×10 <sup>-5</sup>	0.8866±1×10 <sup>-3</sup>	0.8446±1×10 <sup>-3</sup>	0.8576±1×10 <sup>-3</sup>	0.7069±1×10 <sup>-2</sup>
	GraphCDR	0.8076±1×10 <sup>-3</sup>	0.8144±1×10 <sup>-3</sup>	0.7819±1×10 <sup>-3</sup>	0.8107±1×10 <sup>-3</sup>	0.5946±1×10 <sup>-2</sup>
	TSGCNN	0.9156±1×10 <sup>-3</sup>	0.9182±1×10 <sup>-3</sup>	0.8833±1×10 <sup>-3</sup>	0.8906±1×10 <sup>-3</sup>	0.7779±1×10 <sup>-2</sup>
	<b>Ours</b>	<b>0.9623±1×10<sup>-3</sup></b>	<b>0.9773±1×10<sup>-3</sup></b>	<b>0.9032±1×10<sup>-3</sup></b>	<b>0.8966±1×10<sup>-3</sup></b>	<b>0.8133±1×10<sup>-3</sup></b>
CCLE	DeepCDR	0.8322±1×10 <sup>-3</sup>	0.8400±1×10 <sup>-3</sup>	0.8047±1×10 <sup>-3</sup>	0.8246±1×10 <sup>-3</sup>	0.6337±1×10 <sup>-2</sup>
	DeepDSC	0.8497±1×10 <sup>-3</sup>	0.8649±1×10 <sup>-3</sup>	0.8239±1×10 <sup>-3</sup>	0.8373±1×10 <sup>-3</sup>	0.6660±1×10 <sup>-2</sup>
	MOFGCN	0.8442±1×10 <sup>-3</sup>	0.8563±1×10 <sup>-3</sup>	0.8191±1×10 <sup>-3</sup>	0.8344±1×10 <sup>-3</sup>	0.6583±1×10 <sup>-3</sup>
	GraphCDR	0.8438±1×10 <sup>-3</sup>	0.8520±1×10 <sup>-3</sup>	0.8160±1×10 <sup>-3</sup>	0.8317±1×10 <sup>-3</sup>	0.6526±1×10 <sup>-2</sup>
	TSGCNN	0.8823±1×10 <sup>-3</sup>	0.8908±1×10 <sup>-3</sup>	0.8471±1×10 <sup>-3</sup>	0.8585±1×10 <sup>-3</sup>	0.7108±1×10 <sup>-2</sup>
	<b>Ours</b>	<b>0.8925±1×10<sup>-3</sup></b>	<b>0.9564±1×10<sup>-3</sup></b>	<b>0.8443±1×10<sup>-3</sup></b>	<b>0.8568±1×10<sup>-3</sup></b>	<b>0.7963±1×10<sup>-3</sup></b>

<https://doi.org/10.1371/journal.pcbi.1012748.t001>

the GDSC dataset, where the horizontal axis represents the false positive rate and the vertical axis represents the true positive rate; Fig 1B shows the ROC curve of the ASGCL model on the CCLE dataset. According to the figures, the ASGCL model scored 96.23% in AUC on the GDSC dataset and 89.25% in AUC on the CCLE dataset, both higher than the baseline in tests on the two datasets. The independent test results indicate that the ASGCL model has a high generalization ability.

## Predicting task results

Fig 2 visualizes the performance of the ASGCL model in the drug response regression task. The horizontal axis represents the actual IC50 values, and the vertical axis represents the predicted IC50 values, with the color gradient from light to dark indicating increasing density. The solid red line represents the prediction results of linear fitting based on the data; the dashed red line represents the ideal scenario where the predicted results completely match the

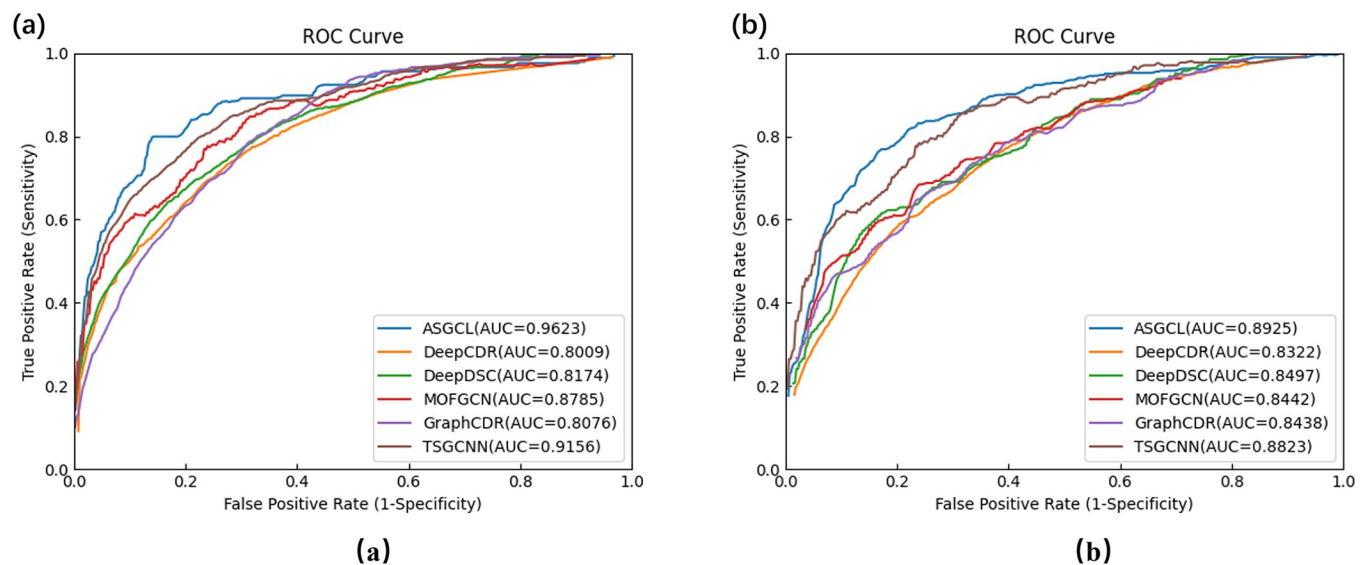
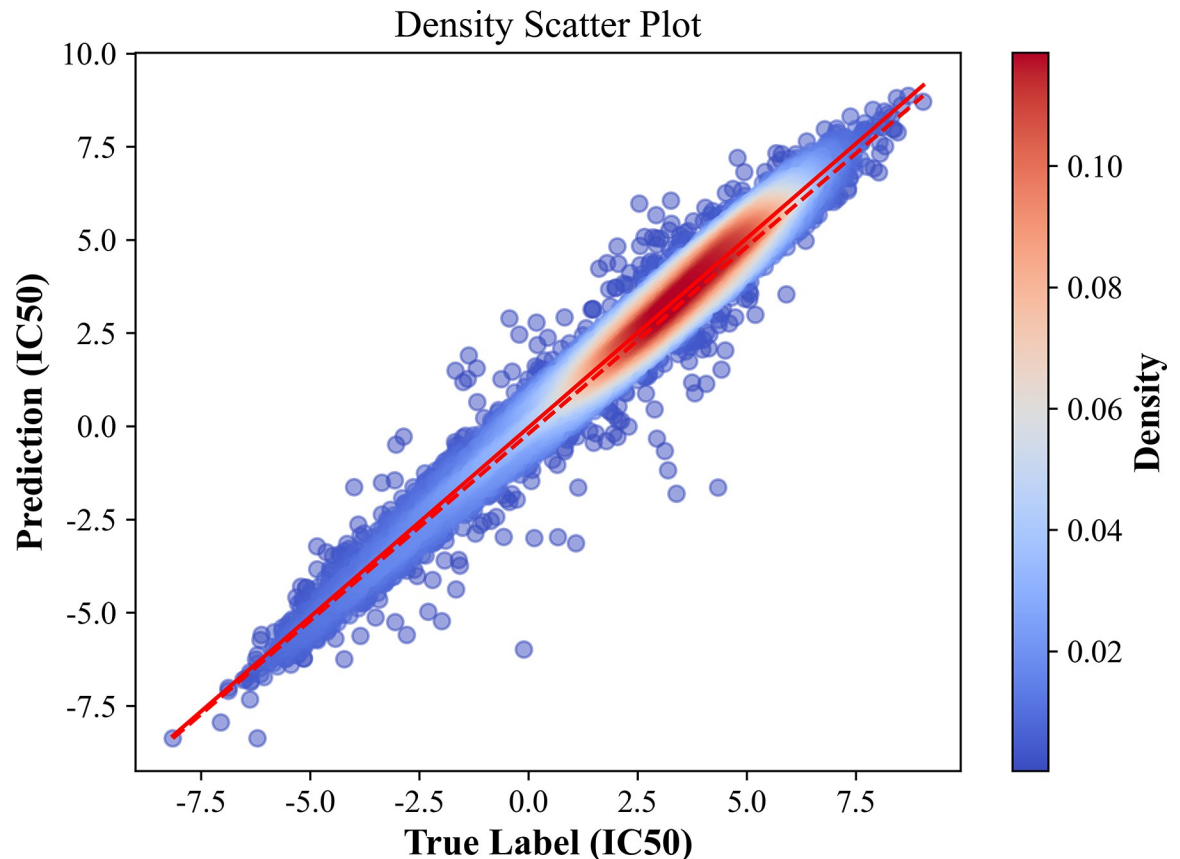


Fig 1. ROC curves of ASGCL model and five comparison models. (a) ROC curves on GDSC dataset (b) ROC curves on CCLE dataset.

<https://doi.org/10.1371/journal.pcbi.1012748.g001>



**Fig 2. Scatter plot of predicting drug response (IC50 value) in GDSC using ASGCL model.**

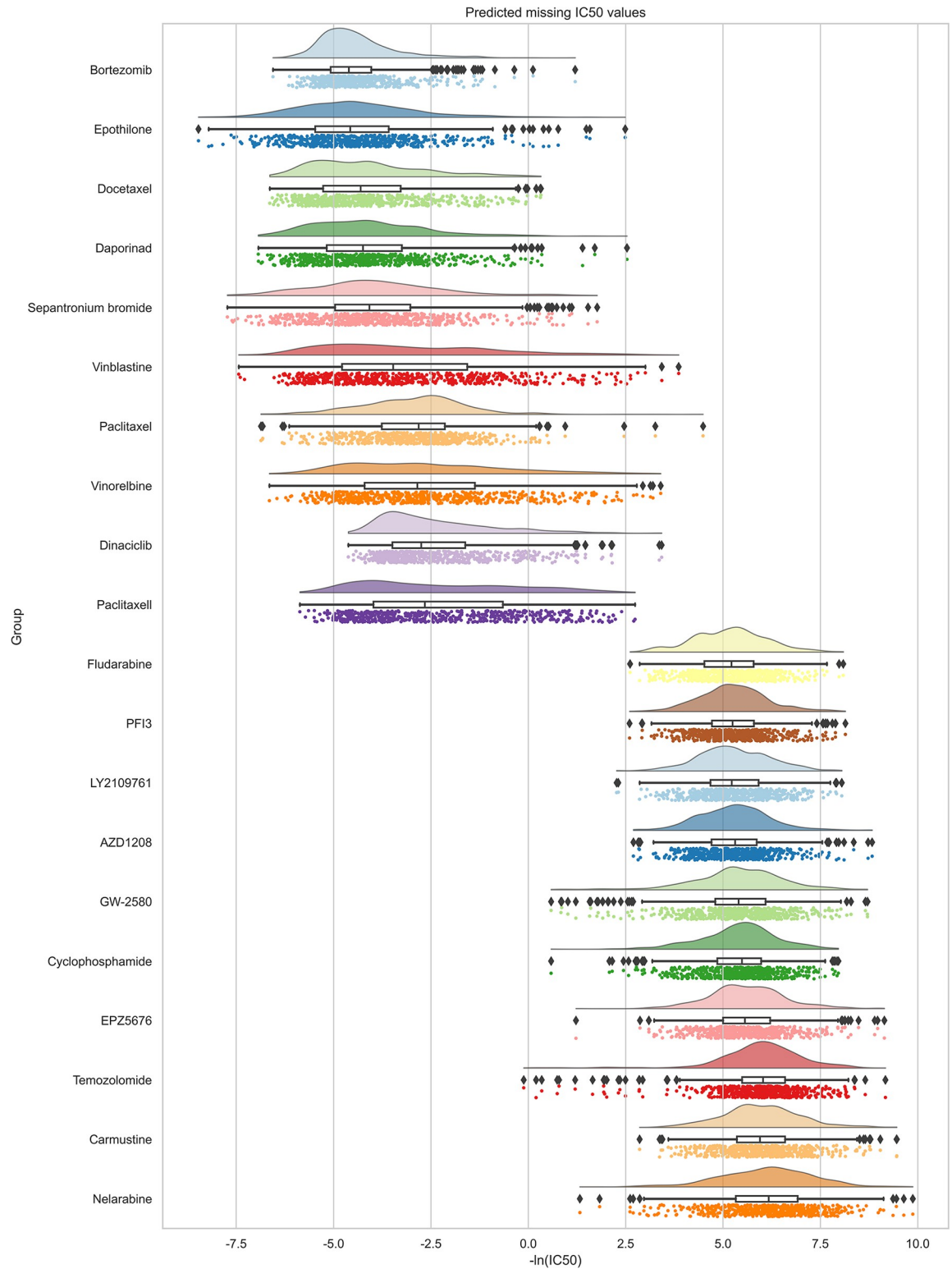
<https://doi.org/10.1371/journal.pcbi.1012748.g002>

actual results. Visualization through scatter plots of the distribution of actual versus predicted values shows a good correlation between actual and predicted responses, demonstrating that the proposed ASGCL model exhibits good predictive performance for a large number of random samples.

### Prediction of response in unknown drug cell lines

We established the ASGCL model using all known cell line-drug reactions and predicted reactions for unknown cell line-drug pairs, as shown in Fig 3.

We applied the ASGCL model to predict unknown cell line-drug reactions in the GDSC database. ASGCL was trained on all known interactions between 962 cell lines and 228 drugs, then predicted missing IC50 values in the GDSC database. Predictions were grouped by drug and then sorted by the median of the predicted IC50 values. The raincloud plots of the top 10 drugs with the highest and lowest median IC50 values are shown in Fig 3. After testing each drug's distribution, more than 17 out of 20 selected drugs showed no significant difference between their IC50 values and the distribution of predicted unknown IC50 values. The results indicate that the predicted missing IC50 values match the distribution of actual measurements, suggesting that our model can effectively distinguish between sensitive and resistant drugs. Since these unknown DCPs lack real values, the results were validated based on relevant literature. Specifically, Bortezomib [51] was predicted to be the most effective drug, consistent with previous studies that suggest Bortezomib can inhibit cancer by interacting with various cancer cell lines.



**Fig 3. Predict the IC50 values of unknown cell line drug reactions grouped by drugs.** Drugs are classified based on the median predicted IC50 values of all missing cell lines, and the top 10 drugs with the highest median IC50 have the worst efficacy; The last 10 drugs with the lowest IC50 median may be the most effective.

<https://doi.org/10.1371/journal.pcbi.1012748.g003>

Table 2. Ablation experiments using different contrasting methods.

Original Graph	Resistance Graph	Node Mask	Link Pruning	AUC	AP	F1_Score
✓				0.9045	0.9448	0.8351
✓	✓			0.9384	0.9712	0.8519
✓	✓	✓		0.9315	0.9615	0.8727
✓	✓		✓	0.9240	0.9626	0.8148
✓	✓	✓	✓	<b>0.9623</b>	<b>0.9773</b>	<b>0.8966</b>

<https://doi.org/10.1371/journal.pcbi.1012748.t002>

## Ablation

This paper conducted ablation experiments on the GDSC dataset for the nonlinear subspace, GraphMorpher module, and graph encoder used in the contrastive learning stage of the ASGCL model, studying the effectiveness of these modules by replacing or eliminating each of them.

## Evaluation of the nonlinear subspace

In this study, we used a nonlinear subspace approach to merge linear and nonlinear features. Although existing linear methods perform well in extracting features of cell lines and drugs, relying solely on linear feature extraction overlooks the topological structure features in cell lines and drugs. Therefore, a nonlinear subspace was introduced to merge linear and nonlinear features, extracting more effective original features.

To evaluate the effect of using a nonlinear subspace in the model, a series of ablation experiments were conducted (Table 2). Comparisons were made between using only linear features, only nonlinear features, and combining both. Fig 4 shows the comparison between the performance of the model using a nonlinear subspace and models that only extract linear or nonlinear features.

The results show that the performance of the model using a nonlinear subspace is significantly better than models that only extract linear or nonlinear features. This indicates that the subsequent contrastive learning process largely depends on effective initial features. By introducing a nonlinear subspace, complex relationships between cell lines and drugs can be better

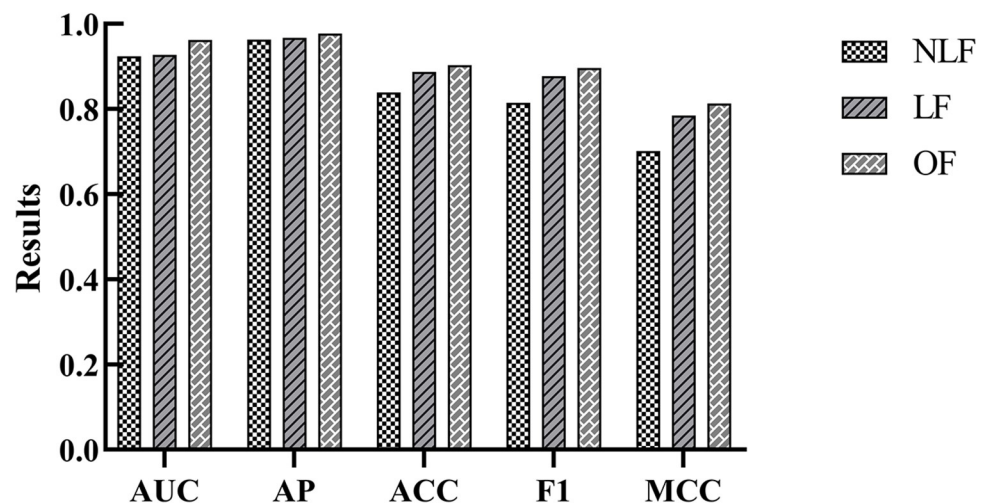


Fig 4. Research on ablation of nonlinear subspaces.

<https://doi.org/10.1371/journal.pcbi.1012748.g004>

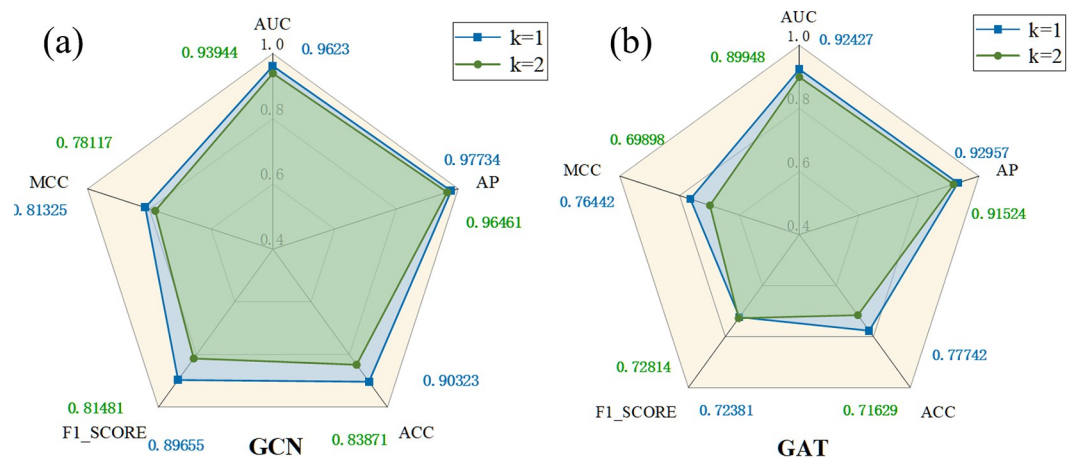
captured, thereby improving the performance of the model. In summary, the method of using a nonlinear subspace in this study can effectively merge linear and nonlinear features and extract more informative original features.

### Evaluation of contrastive learning

This study introduces contrastive learning to enhance the quality of feature extraction. Although supervised learning has been effective, the addition of contrastive learning enables the model to more effectively identify and extract key information in the graph, thus obtaining more distinctive feature representations. To verify the role of contrastive learning in the ASGCL model, this paper conducted ablation experiments to demonstrate the effectiveness of the proposed method in the following five parts: (1) Directly performing drug resistance prediction tasks on the features of the original graph, without contrastive learning; (2) Only conducting contrastive learning on sensitive and resistant graphs, without learning nodes and topological structures; (3) Only performing node masking operations on the input graph structure; (4) Only performing link pruning operations on the input graph structure; (5) Performing node masking and link pruning operations separately on the input sensitive and resistant graphs.

### Evaluation of the graph encoder

In this study, a graph encoder was used to encode the node attributes and their topological structure within the graph structure. During the model construction process, various commonly used encoders were considered, including Graph Convolutional Network (GCN) and Graph Attention Network (GAT). This paper investigated the impact of different encoders on the overall performance of the model through ablation experiments, as well as the effect of setting different convolutional layers in the graph encoder, with comparative results detailed in Fig 5. The experimental results showed that using GCN as the graph encoder with  $K = 1$  yielded better overall effects than GAT, with an increase of 3.81% in AUC, and other performance metrics such as AP and F1 Score improved by 4.78% and 17.27% respectively. This result indicates that GCN has a significant advantage in encoding graph structures in this study. This superiority might be attributed to GCN's ability to effectively capture key structural features of the graph while maintaining a simple structure. In contrast, the performance



**Fig 5. Comparative experiments on different graph encoders in different neighborhoods.** (a) Experimental results of GCN on different neighborhood layers (b) Experimental results of GAT on different neighborhood layers

<https://doi.org/10.1371/journal.pcbi.1012748.g005>

decline of GAT could be due to overfitting, as the specificity of cell line and drug data may not suit the complex attention mechanism.

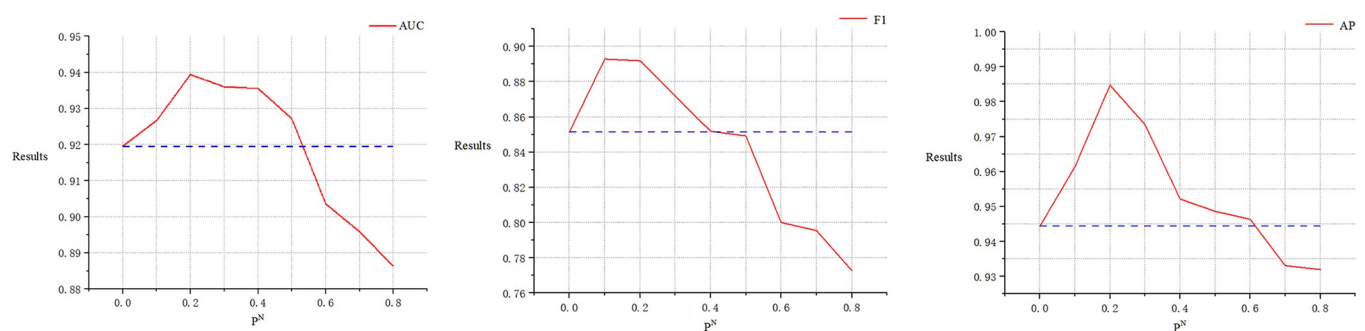
The impact of the number of convolutional layers  $K$  on the model's performance, as shown in Fig 5, indicates a decline when  $K$  increases from 1 to 2. This suggests that while a higher number of neighborhood layers may provide more valuable information, the specificity of the cell line and drug data in this study means that too high a value of  $K$  could introduce excessive noise or irrelevant information. Furthermore, the issue of over-smoothing could arise, leading to generated feature representations lacking distinctiveness, thereby negatively impacting model performance.

## Parameter experiments

In this section, we study the impact of several key parameters on the model. While assessing any parameter, default values are set for the others. The following parameter studies are conducted on the GDSC dataset.

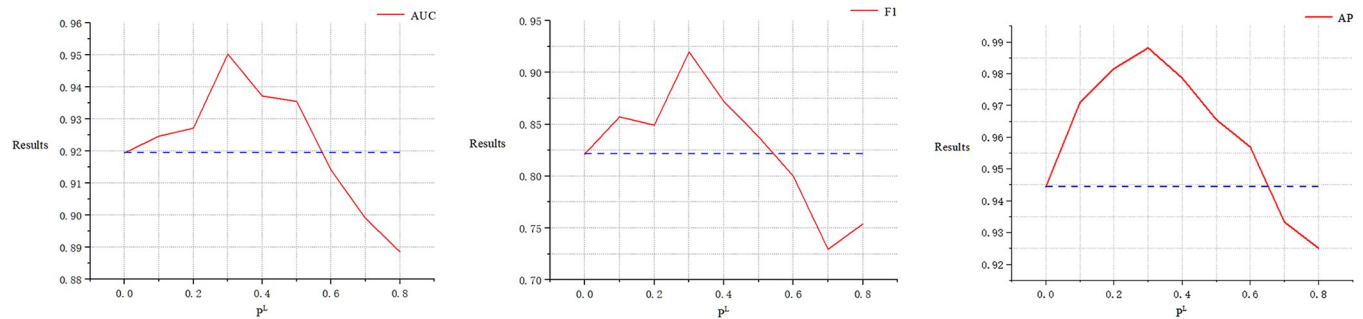
### The roles of node masking probability $\rho^N$ and link pruning probability $\rho^L$ in GraphMorpher

Figs 6 and 7 analyze the impact of two different hyperparameters on segmentation performance. In the proposed GraphMorpher,  $\rho^N$  represents the probability of node attributes in the graph structure being masked, and  $\rho^L$  represents the probability of the topological structure being pruned. To study the impact of these two probabilities on experimental results, comparative experiments on these parameters were conducted in the dataset. The values of the two parameters were incrementally increased from 0.1 to 0.8 for experimentation, with a step size of 0.1. When experimenting with any one parameter, the other parameter was set to 0 to ensure that the remaining parameters did not affect the experiment. The results indicate that as the parameters  $\rho^N$  and  $\rho^L$  gradually increased, the effectiveness of the experiment first improved and then declined. Specifically, the best experimental effect was achieved when the node masking probability  $\rho^N$  was at 0.2 and the topology pruning probability  $\rho^L$  was at 0.3. As shown in the figure, the model's effectiveness decreased as the values of  $\rho^N$  and  $\rho^L$  exceeded certain thresholds, likely due to the excessive probability of damaging nodes and links, causing the graph's underlying semantic information and topological structure to suffer devastating damage, leading to feature loss. Based on the experimental results, we set the parameters  $\rho^N$  and  $\rho^L$  to 0.2 and 0.3 respectively in the experiments, and set the truncation probabilities  $\mathcal{T}_N$



**Fig 6. The impact of the parameter  $\rho^N$  on the ASGCL model's performance is demonstrated, where the blue dotted line represents the baseline result of the ASGCL model on the GDSC dataset when the parameter  $\rho^L$  is not introduced.**

<https://doi.org/10.1371/journal.pcbi.1012748.g006>



**Fig 7. The impact of the parameter on  $\rho^L$ .** The ASGCL model's performance is demonstrated, where the blue dotted line represents the baseline result of the ASGCL model on the GDSC dataset when the parameter  $\rho^N$  is not introduced.

<https://doi.org/10.1371/journal.pcbi.1012748.g007>

and  $\mathcal{T}_L$  to 0.3 and 0.4, respectively, to prevent excessive probability of graph structure sparsification, ensuring the graph structure is not severely damaged.

### The combined role of node masking probability $\rho^N$ and link pruning probability $\rho^L$ in GraphMorpher

As shown in Fig 8, the model exhibited different effects under the combined action of parameters  $\rho^N$  and  $\rho^L$ . The experimental results indicate that the optimal choices for these two parameters are 0.2 and 0.3, respectively, under which the model achieved the highest AUC.

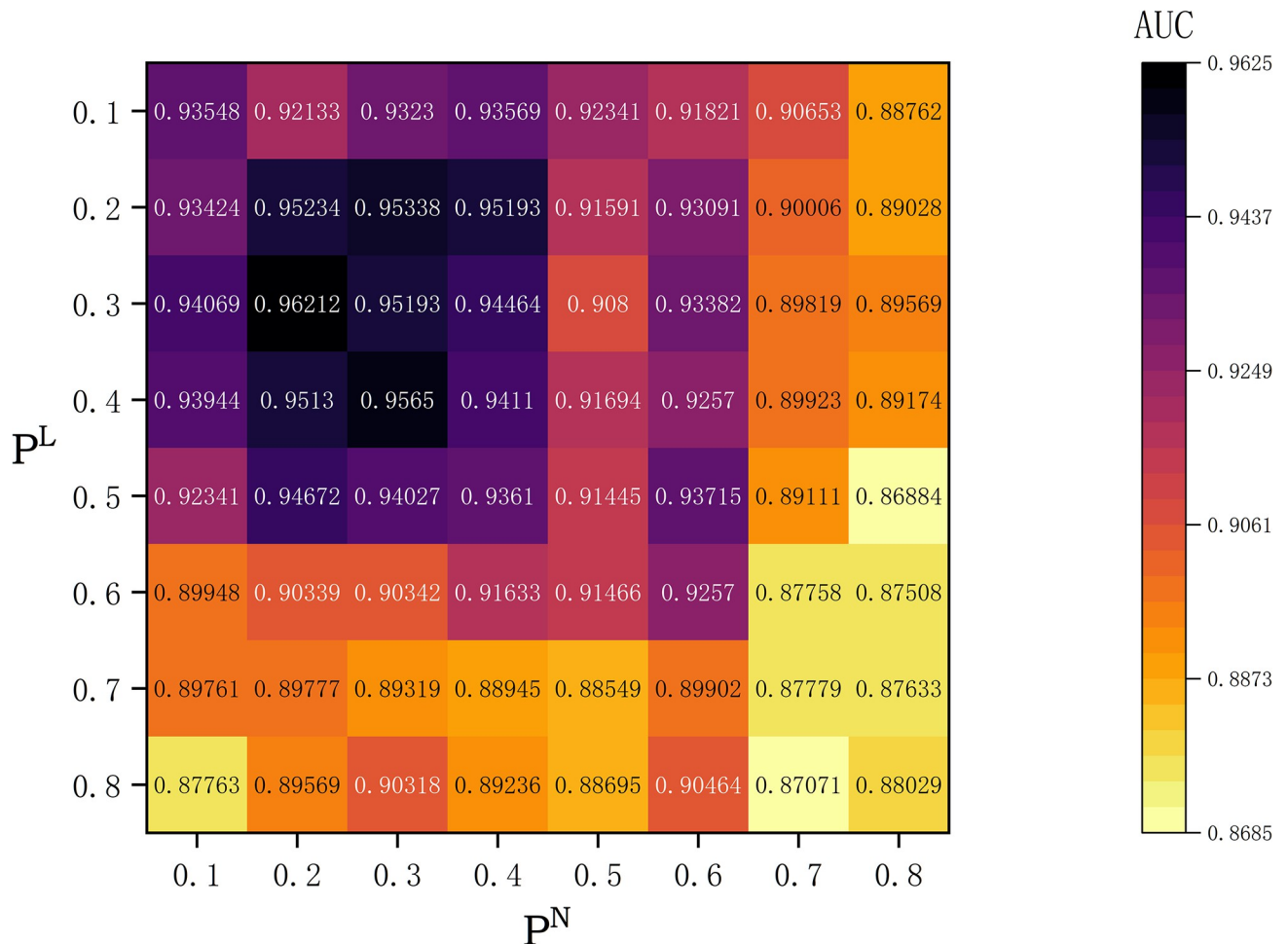
Furthermore, Fig 8 demonstrates that an appropriate sparsity ratio can help the model remove redundant noise and enhance its learning ability, while an overly high sparsity ratio will damage the feature attributes and topological structure of the graph, reducing the model's performance, thereby verifying the robustness and effectiveness of the GraphMorpher module.

## Conclusion

The main work of this article is to propose the ASGCL model, which utilizes contrastive learning and graph neural networks to extract features of cells and drugs and enhance the model's discriminative ability for predicting drug resistance. This study conducted a large number of experiments, and compared with benchmark methods on actual datasets, the ASGCL model showed higher efficiency in predicting drug resistance tasks. In addition, the ASGCL model adopts an end-to-end training method to enrich the learned feature information. Overall, the experiment verified the reliability of the ASGCL model, which can accurately predict drug resistance tasks on sparse data. This indicates that end-to-end graph contrastive learning methods can help improve the predictive ability for drug resistance tasks. However, the ASGCL model still has some limitations. Future work will focus on finding more effective and universal graph data augmentation methods to further improve the performance and applicability of models.

## Discussion

Predicting cancer drug response is the main goal of using neural networks in the field of machine learning to assist cancer drug discovery and precision medicine in cancer research. In this work, we developed a model called ASGCL that utilizes multiple levels of contrastive learning tasks to enhance the model's learning ability, providing a new deep learning framework for exploring the interactions between drugs and cell lines. The predictive ability of ASGCL has been validated on the GDSC dataset and compared with state-of-the-art methods under



**Fig 8. The parameter  $\rho^N$  and  $\rho^L$  impact the performance of the ASGCL model on the GDSC dataset.** In the graph, the colors represent the magnitude of the ASGCL's performance: darker colors correspond to higher AUC values, while lighter colors indicate lower AUC values.

<https://doi.org/10.1371/journal.pcbi.1012748.g008>

various evaluation settings. Extensive case studies have demonstrated the ability of ASGCL to interpret and predict cell lines and drug responses, as well as to search for relevant response outcomes.

Despite our efforts, there is still room for improvement. The performance of ASGCL in predicting missing drug and cell line CDRs was lower than expected. In such problems, due to the sparsity of the data type itself and the lack of prior knowledge of drugs/cell lines in the model, ASGCL can only fit the reaction process of missing data through existing data. Taking inspiration from prior knowledge in the medical field, introducing some regularization processes or knowledge graphs as prior knowledge into the model may greatly alleviate this problem. In addition, due to the inherent non interpretability of deep learning, it is expected that in future research, more mathematical methods or data refinement will be used to determine the role of models in predicting cell line drug interactions.

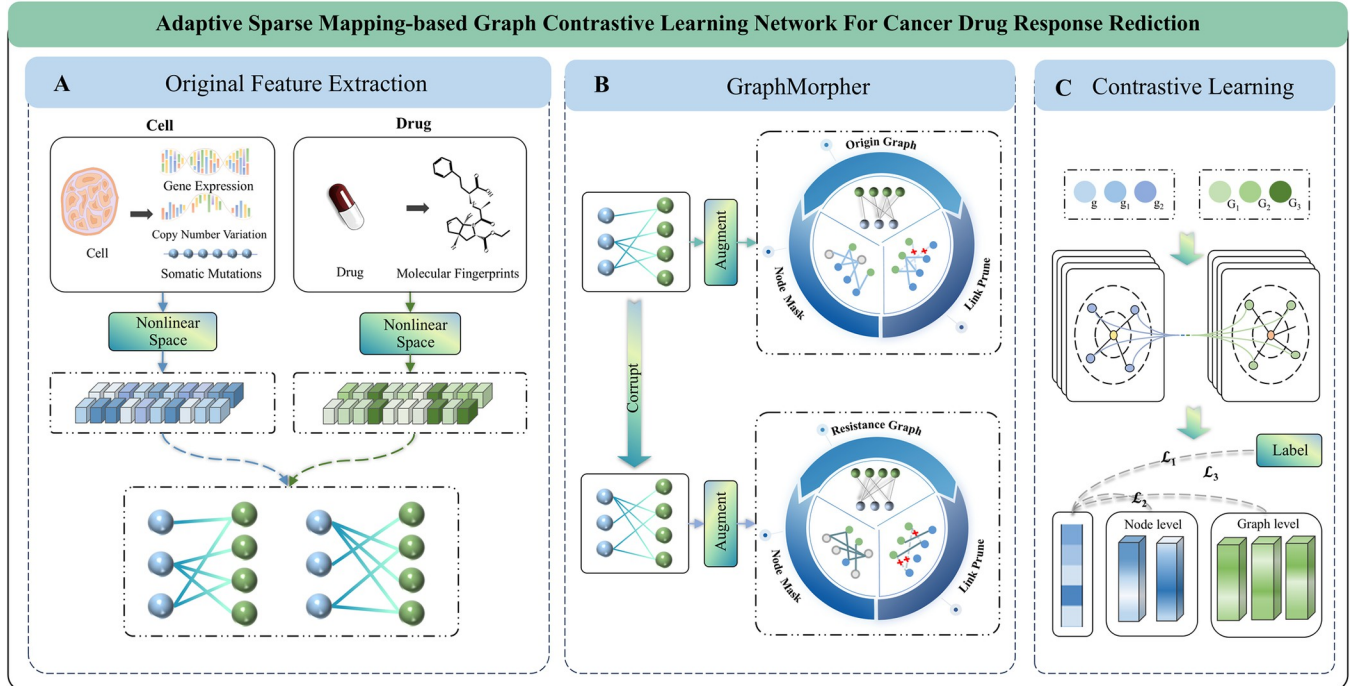
## Method

Now we introduce the ASGCL model and then describe its workflow in detail. The paper presents a method named the ASGCL model, aimed at predicting the degree of response between

cell lines and drugs. The ASGCL model integrates linear and nonlinear features of cell lines with drug characteristics through utilizing nonlinear subspaces. It employs contrastive learning to extract efficient representations of interactions between cell lines and drugs, thereby predicting tasks related to drug resistance.

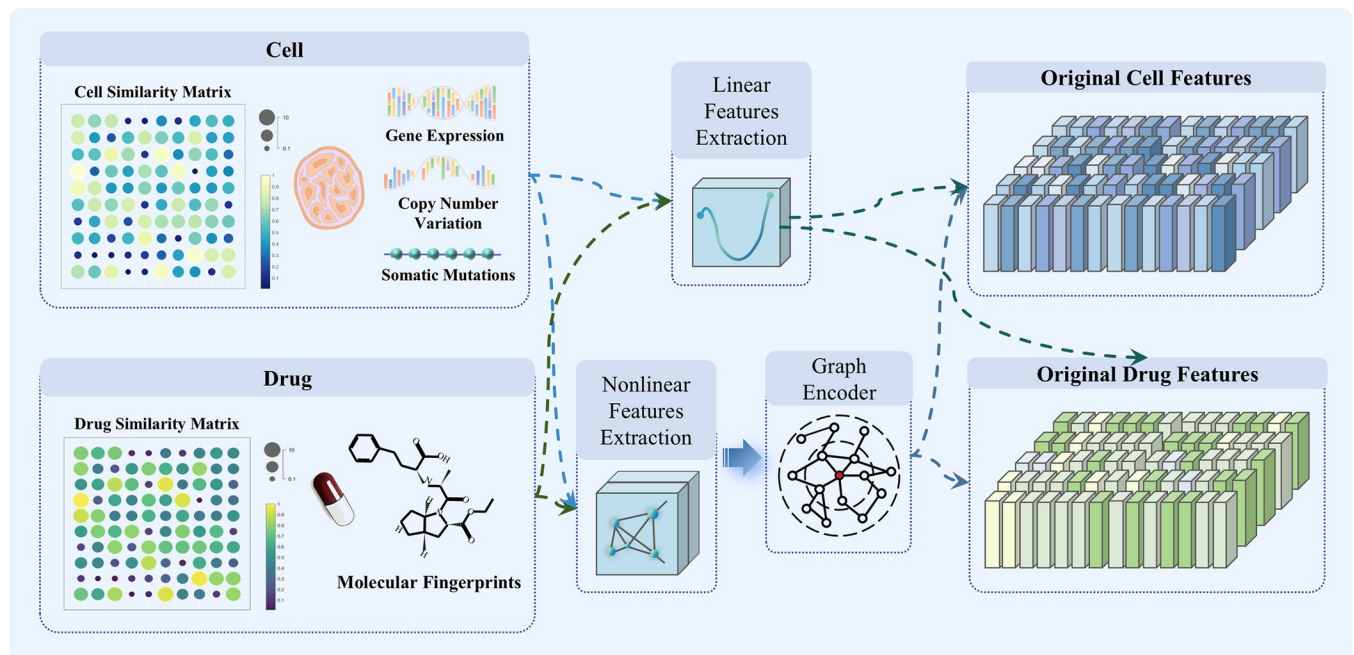
**Overview**

The ASGCL model (Fig 9) initially constructs a nonlinear subspace to extract and integrate the linear and nonlinear features of cell lines and drugs, forming their primary characteristics. Subsequently, these features are mapped to various nodes of a heterogeneous graph, using the relationships between nodes to represent interactions between cell lines and drugs. Next, we design the GraphMorpher module to enhance the graph structure’s expressive capability, which transforms the original graph into multiple enhanced graphs by augmenting node attributes and topological structures within the graph. These enhanced graphs are then processed by a graph encoder, generating feature representations rich in semantic information. To further enhance the model’s feature expressiveness and generalization, we employ a contrastive learning strategy, defining positive and negative samples. This approach yields more discriminative feature representations through the comparison of positive and negative samples, thereby enhancing predictions for drug resistance tasks. Ultimately, the model optimization merges supervised loss with contrastive loss, where supervised loss guides the learning of task-specific knowledge, and contrastive loss facilitates the learning of diverse feature representations. Through the joint optimization of these two types of loss, a more accurate prediction of cell line and drug reactions is achieved.



**Fig 9. Schematic diagram of the ASGCL model.** Module A utilizes a nonlinear subspace to extract cell line and drug features as primary characteristics; Module B, named GraphMorpher, adaptively sparsify the input graph structure; Module C is a contrastive learning module, which enhances the model’s discriminative ability by processing and comparing multiple graph structures.

<https://doi.org/10.1371/journal.pcbi.1012748.g009>



**Fig 10. Schematic diagram of nonlinear subspace module.**

<https://doi.org/10.1371/journal.pcbi.1012748.g010>

## Nonlinear subspace

This study aims to develop the ASGCL model to learn feature representations of cell lines and drugs, and to predict the degree of their interaction. To achieve this objective, the paper extracts features from two perspectives: linear relationships and nonlinear topological structures, as shown in Fig 10. Initially, linear features of cell lines and drugs are extracted by processing their multi-omics data through linear functions. Subsequently, the features of cell lines and drugs are mapped onto graph structures. Using graph encoders, the nonlinear topological structures of cell lines and drugs are extracted, learning the linkages between nodes in the graph to represent their nonlinear features. Ultimately, the linear and nonlinear features are integrated to form a comprehensive feature representation of cell lines and drugs. By combining linear and nonlinear features, this approach not only captures the complexity of cell line and drug data but also provides a more complete and rich feature representation. These features serve as inputs for the GraphMorpher module, aiming to enhance the accuracy of predictions for cell line-drug reactions.

## Linear features

The model inputs are multi-omics features of cell lines and molecular fingerprints of drugs. Matrix  $A \in \mathbb{R}^{m \times n}$  is used as the cell line-drug association matrix, where  $m$  denotes the number of cell lines and  $n$  the number of drugs.  $A_{ij} = 1$  indicates that drug  $j$  is sensitive to cell line  $i$ ; otherwise,  $A_{ij} = 0$ .

Since cell line gene expression and drug molecular fingerprint features usually have different dimensions, weight matrices  $W_c$  and  $W_d$  are used for linear transformation of cell line gene expression and drug expression, as shown in Eq 8:

$$E_c = C_i \times W_c \quad (8)$$

where  $E_c$  represents the characteristics of the cell line, and  $C_i$  represents the gene expression

value of the  $i$ -th cell line in the gene expression matrix.

$$E_d = D_i \times W_d \quad (9)$$

where  $E_d$  represents the feature representation of the drug, and  $D_i$  denotes the expression value of the  $i$ -th drug in the drug molecular fingerprint matrix.

After obtaining the feature representations of cell lines and drugs, the features are normalized by calculating the normalization coefficients, as shown in Eq 10 and Eq 11:

$$N_C = \left( \sum_j A_{ij} + 1 \right)^{-1} + I_m \quad (10)$$

$$N_D = \left( \sum_j A_{ji} + 1 \right)^{-1} + I_n \quad (11)$$

where  $N_C$  and  $N_D$  respectively represent the normalized feature coefficients of cell lines and drugs,  $I_m$  is an  $m$ -dimensional identity matrix, and  $I_n$  is an  $n$ -dimensional identity matrix.

In summary, this paper employs linear functions to transform the original node features of cell lines and drugs into their linear features, as depicted in Eq 12 and Eq 13

$$LF_{Cell} = \omega_1 \cdot E_c \cdot N_C \quad (12)$$

$$LF_{Drug} = \omega_2 \cdot E_D \cdot N_D \quad (13)$$

where  $\omega_1, \omega_2$  is a learnable weight parameter.

## Nonlinear features

This study assumes that similar drugs have similar reactions to similar cell lines. Considering network structure and feature attributes, the study maps cell lines and drugs onto an isomorphic graph, learning node similarities through a graph encoder to infer their reactions.

Cell lines are represented through a cell line similarity matrix, with the constructed graph  $G_C = (X_c, A_C)$  representing cell lines. In the graph  $G_C$ , nodes represent cell lines and links between nodes indicate similarities among them.  $X_c$  is the feature representation of cell lines, recording the characteristics of all genes in the cell line;  $A_C \in \mathbb{R}^{N_c \times N_c}$  is the adjacency matrix of  $G_C$ , indicating the degree of similarity among cell lines, where  $N_c$  is the number of cell lines.

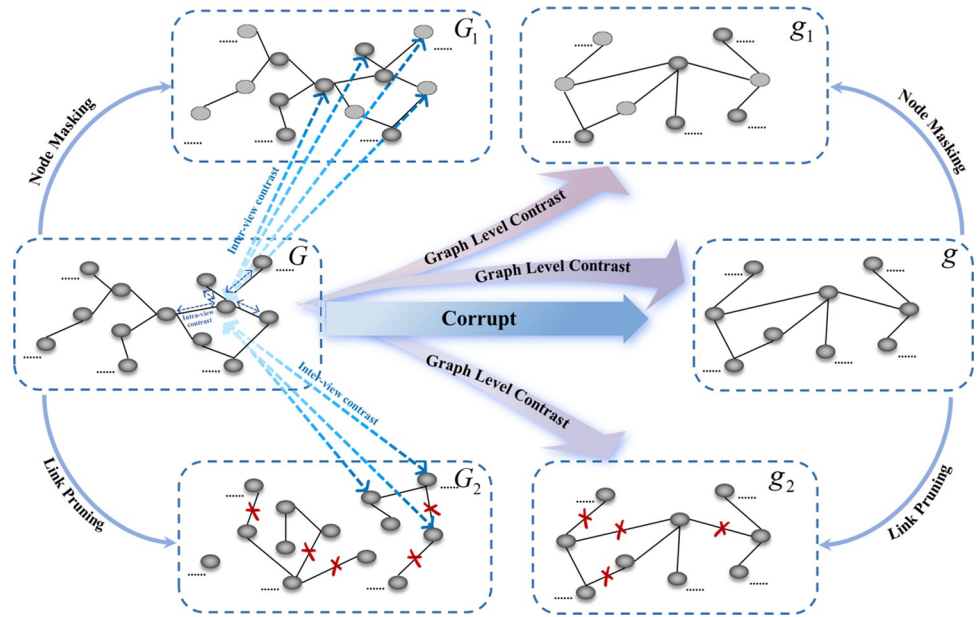
Drugs are represented through a drug similarity matrix, with the constructed graph  $G_D = (X_D, A_D)$  representing drugs. In the graph  $G_D$ , nodes represent drugs and links indicate similarities between them.  $X_D$  is the feature representation of drugs, recording the characteristics of all drugs;  $A_D \in \mathbb{R}^{N_d \times N_d}$  is the adjacency matrix of  $G_D$ , indicating the similarity among drugs, where  $N_d$  is the total number of drugs.

Using the above methods, this study employs a graph neural network encoder to encode node features and their similarities in the graph, achieving effective propagation of feature information among related nodes. During this process, it captures the characteristics of cells and drugs and reveals the latent nonlinear feature representations in cell lines and drugs through the information propagation mechanism. This approach enables the study to more deeply understand and predict the complex interactions between cells and drugs.

After obtaining the linear and nonlinear features of cell lines and drugs, they are combined to obtain the final representation of cell lines and drugs, as shown in Eq 14 and Eq 15.

$$X_C = LF_{Cell} + NLF_{Cell} \quad (14)$$

$$X_D = LF_{Drug} + NLF_{Drug} \quad (15)$$



**Fig 11. GraphMorpher diagram.**

<https://doi.org/10.1371/journal.pcbi.1012748.g011>

In summary, after integrating the linear and nonlinear features of cell lines/drugs through a nonlinear subspace, based on the sensitive response between cell lines and drugs, an undirected heterogeneous graph  $G_{Sen} = (X_{CD}, A_{Sen})$  can be constructed. Here,  $X_{CD} = \begin{bmatrix} X_C \\ X_D \end{bmatrix}$  represents the sum of the cell line feature representation  $X_C$  and the drug feature representation  $X_D$ ;  $A_{Sen}$  is the adjacency matrix of  $G_{Sen}$ , representing the reaction relationship between cell lines and drugs, where  $A_{Sen} = 1$  indicates sensitivity of the cell line to the drug, and  $A_{Sen} = 0$  otherwise.

### GraphMorpher

Based on the overall framework of contrastive learning [28,29,30,44,45,52], a GraphMorpher module is proposed. The primary function of GraphMorpher is to generate multiple enhanced views for the original graph  $G_{Sen}$ , which then aids in better feature extraction of  $G_{Sen}$  within the contrastive learning module, as shown in Fig 11 and detailed in Algorithm 1:

**Algorithm 1:** GraphMorpher

**Input:** Original Graph:  $G = (X, A)$ ; Mask probability:  $\rho^N$ ; Pruning probability:  $\rho^L$

**Output:**  $G; G^N; G^L; \tilde{G}; \tilde{G}^N; \tilde{G}^L$

- 1: */\* \*\* Structure new heterogeneous graphs  $\tilde{G}$*
- 2: Calculate  $\tilde{A}$  via Eq 16
- 3:  $\tilde{G} \leftarrow (X, \tilde{A})$
- 4: for  $G = (X, A)$  and  $\tilde{G} = (X, \tilde{A})$  do
- 5: */\* \*\* Calculate sparsity probability*
- 6: Calculate  $P_{Mask}$  via Eq 17 and Eq 18
- 7: Calculate  $P_{prune}$  via Eq 21 and Eq 22
- 8: */\* \*\* Sampling masked nodes*
- 9: Calculate  $\mathbf{X}'$  via Eq 20

```

10:  /**** Sampling pruned links
11:  Calculate  $\mathbf{A}'$  via Eq 23
12:  /****Construct augment graph
13:   $\mathbf{G}^N \leftarrow (X', A)$ 
14:   $\mathbf{G}^L \leftarrow (X, A')$ 
15:  end for
16:  Return  $G; G^N; G^L; \tilde{G}; \tilde{G}^N; \tilde{G}^L$ 

```

Inspired by the definition of positive and negative samples in contrastive learning, this paper inverts the adjacency matrix  $A_{Sen}$  of the original graph  $G_{sen}$  to obtain a new adjacency matrix  $A_{Res}$ , as shown in Eq 16:

$$A_{ij}^{Res} = \begin{cases} 0, A_{ij}^{Sen} = 1 \\ 1, A_{ij}^{Sen} = 0 \end{cases} \quad (16)$$

The adjacency matrix  $A_{Sen}$  represents the sensitivity between cell lines and drugs; thus, the new adjacency matrix  $A_{Res}$  represents the resistance response between them. Keeping the node features  $X_{CD}$  unchanged and using the new adjacency matrix  $A_{Res}$ , a new resistance graph  $G_{Res} = (X_{CD}, A_{Res})$  is formed.

It is believed that using completely opposite links to construct graph  $G_{Res}$  allows  $G_{sen}$  to improve its learned features by maximizing the distance to  $G_{Res}$ , thereby enhancing the model's discriminative ability.

Upon obtaining graph  $G_{Res}$ , an adaptive graph augmentation method was designed based on node attributes and topological structure in the graph. Specifically, for any given graph structure, two new views are generated by adaptively masking original node attributes and pruning the topological structure; then, contrastive loss is computed to maximize the consistency of node representations in these two views; finally, through contrastive learning between nodes, the final node representation is obtained for predicting the degree of reaction between cell lines and drugs. This method enables the model to better learn node features, forcing it to recognize the semantic information of nodes; moreover, it prunes links based on the graph's topological structure, thereby highlighting important link structures.

## Node mask

At the node attribute level, this study adopts an adaptive approach to mask the attributes of nodes in the graph, by masking nodes with sparse information and emphasizing nodes with rich information, thus better extracting underlying semantic information. In contrast, many previous studies [52,53] have destructively pruned nodes, damaging not only the underlying semantic features of the nodes but also affecting the topological structure between connected nodes. Particularly in drug resistance prediction tasks, due to the inherent sparsity of medical data, such destructive methods further sparsify the data, preventing the extraction of key features.

In the process of calculating node masks, considering the amount of information contained in nodes,  $\mathcal{F}(v)$  is defined as the centrality of the feature vector of node  $v$  [54,55].  $\mathcal{F}(v)$  takes into account the importance of neighboring nodes, thereby assigning different contributions to each neighbor.  $\mathcal{F}(v)$  is calculated as the eigenvector corresponding to the largest eigenvalue of its adjacency matrix, representing the most contributing factor. The  $\mathcal{F}(v)$  of each node is proportional to the sum of  $\mathcal{F}(v)$  of its neighboring nodes, and nodes connected to many neighbors or to nodes with greater influence will have a higher  $\mathcal{F}(v)$ . In this paper, features that frequently appear in nodes are considered important, thus a centrality measure  $\mathcal{M}_i^N$  is

defined for each node  $v$ , and  $\mathcal{M}_i^N$  is used as the measure of node  $v$ 's features, as shown in Eq 17:

$$\mathcal{M}_i^N = \sum_{v \in \mathcal{V}} |x_v^i| \cdot \mathcal{F}(v) \tag{17}$$

where  $|x_v^i|$  representing the feature size of node  $v$  in dimension  $\mathcal{F}(v)$  represents the contribution of the node. Next, the probability  $\mathcal{P}_{Mask}$  of a node being masked is obtained through normalization using weights, where  $\mathcal{P}_{Mask}$  indicates the importance of the node, as shown in Eq 18.

$$\mathcal{P}_{Mask} = \min \left( \rho^N \cdot \frac{\mathcal{M}_{max}^N - \mathcal{M}_i^N}{\mathcal{M}_{max}^N - \mu_{\mathcal{M}}^N}, T_N \right) \tag{18}$$

where  $\rho^N$  controls the probability of node masking,  $\mathcal{M}_{max}^N$  and  $\mu_{\mathcal{M}}^N$  are the maximum and average values of  $\mathcal{M}_i^N$ , respectively, and  $T_N$  is the truncation probability of node masking, used to prevent excessively high masking probabilities, as overly high probabilities would lead to excessive sparsity of semantic features in the graph, making it impossible to extract effective semantic features in subsequent encoding processes.

After obtaining the masking probability  $\mathcal{P}_{Mask}$  of nodes, a portion of node dimensions and some feature information are adaptively masked according to  $\mathcal{P}_{Mask}$ .

In the adaptive masking phase, sampling is first performed on a random vector  $m \in \{0,1\}$ , where each dimension of  $m$  is independently sampled according to  $\mathcal{P}_{Mask}$  from a Bernoulli distribution, as shown in Eq 19.

$$m_i \sim \text{Bern}(1 - \mathcal{P}_{Mask}) \tag{19}$$

Subsequently, the generated node feature  $X$  is shown in Eq 20:

$$X = x_1 \cdot m_i; x_2 \cdot m_i; \dots \dots \dots x_N \cdot m_i \tag{20}$$

where operators  $(;)$  and  $(\dots \dots \dots)$  represent connection operators, and  $(\cdot)$  represents element multiplication operations.

### Link pruning

In processing the graph structure, this study employs an adaptive pruning method to enhance the graph structure. By setting different pruning probabilities, the graph structure is refined, removing links with less information and emphasizing key links, thereby optimizing the information density and analysis process of the graph, focusing more on key features. This approach not only enhances the information density of the graph structure but also optimizes the subsequent graph analysis process, enabling more effective capture and utilization of key information in the graph, thus significantly improving the predictive performance of the model.

For each link, this paper defines a link centrality measure  $\mathcal{M}_{ij}^L$ .  $\mathcal{M}_{ij}^L$  is based on the node centrality measures  $\mathcal{M}_i^N$  and  $\mathcal{M}_j^N$  of the nodes at both ends of the link  $(i,j)$ , defining its degree of influence. A logarithmic function is applied to mitigate the impact of nodes with highly dense links, as shown in Eq 21:

$$\mathcal{M}_{ij}^L = \log \left( \frac{\mathcal{M}_i^N + \mathcal{M}_j^N}{2} \right) \tag{21}$$

The probability of pruning links in the graph,  $\mathcal{P}_{prune}$ , is adaptively adjusted based on the  $\mathcal{M}_{ij}^{\mathcal{L}}$  of each edge, thus maintaining the integrity of the graph's topological structure, as shown in Eq 22:

$$\mathcal{P}_{prune} = \min\left(\rho^{\mathcal{L}} \cdot \frac{\mathcal{M}_{max}^{\mathcal{L}} - \mathcal{M}_{ij}^{\mathcal{L}}}{\mathcal{M}_{max}^{\mathcal{L}} - \mu_{\mathcal{M}}^{\mathcal{L}}}, \mathcal{T}_{\mathcal{L}}\right) \tag{22}$$

where  $\rho^{\mathcal{L}}$  controls the probability of link pruning,  $\mathcal{M}_{max}^{\mathcal{L}}$  and  $\mu_{\mathcal{M}}^{\mathcal{L}}$  are the maximum and mean values of  $\mathcal{M}_{ij}^{\mathcal{L}}$ , respectively, and  $\mathcal{T}_{\mathcal{L}}$  is the truncation probability of the link, used to prevent excessively high probabilities of pruning, ensuring that the graph structure does not suffer from destructive information loss. Formally, based on the pruning probability  $\mathcal{P}_{prune}$ , pruning operations are performed in the original set of links  $\mathcal{L}$ , sampling a pruned subset  $\mathcal{L}_{sub}$ , as shown in Eq 23:

$$P\{(i,j) \in \mathcal{L}_{sub}\} = 1 - \mathcal{P}_{prune} \tag{23}$$

where  $P\{(i,j) \in \mathcal{L}_{sub}\}$  is the probability that the link between nodes  $i$  and  $j$  is retained. The subset  $\mathcal{L}_{sub}$  represents the collection of remaining links after pruning the original set  $\mathcal{L}$ .  $\mathcal{P}_{prune}$  is the pruning probability for edges, reflecting the importance of links. The more important the link, the lower the probability it will be pruned. This pruning operation is more likely to prune unimportant links, while keeping important link structures intact in the enhanced view.

By using the aforementioned adaptive pruning method, the input graph structure can be sparsified, focusing on reducing graph structures with less information while preserving key topological structures in the graph, thereby highlighting important structural features in the graph. In contrastive learning, the adaptively sparsified graph structure helps the original graph to learn and reinforce key topological structure features more efficiently. This strategy not only improves the efficiency and accuracy of feature extraction but also provides a more solid foundation for deeply understanding and predicting complex network behaviors.

Finally, through the proposed node masking and link pruning operations, this paper generates two updated views based on the node features and topological structure of the input graph. In experiments, different masking probabilities  $\rho^N$  and pruning probabilities  $\rho^{\mathcal{L}}$  are set to provide the graph encoder with diverse contextual semantic features. The GraphMorpher module randomly and adaptively generates enhanced graphs in each iteration, maintaining important structures and attributes while perturbing potentially unimportant features and topological structures. In this paper, the graph generated from  $G_{sen}$  through node masking is denoted as  $G_{Sen}^N = (X_{CD}^M, A_{Sen})$ , and the graph generated through link pruning is denoted as  $G_{Sen}^L = (X_{CD}, A_{Sen}^P)$ ; similarly,  $G_{Res}$  generates  $G_{Res}^N = (X_{CD}^M, A_{Res})$  and  $G_{Res}^L = (X_{CD}, A_{Res}^P)$ .

### Contrastive learning

This study focuses on contrasting node attributes and link structures within the graph structure, utilizing positive and negative sample analysis to learn feature representations, constructing a multi-relational link model for drug resistance prediction, and employing a graph encoder to extract features of cell lines and drugs. In the graph structure model, neighbor nodes of cell lines or drugs propagate different information in various interaction environments. Therefore, in determining the network structure, different information propagation paths were considered, including various types of encoders and the scope of K-order neighborhoods. By effectively utilizing information sharing among nodes in the graph and aggregation operations in neighborhoods, the information of neighbor nodes is aggregated, thus obtaining

the final node embeddings. Referring to the definition of positive and negative samples in contrastive learning, samples of different levels were set up in the experiments. At the node level, node-level contrastive learning is conducted by comparing the original graph and the graph processed by GraphMorpher; similarly, at the graph level, graph-level contrastive learning is conducted by comparing the original graph and the resistance graph processed by GraphMorpher. In node-level contrastive learning, positive and negative samples for each node of a given graph are defined, where the same nodes in other graph structures are considered as positive samples, and different nodes as negative samples. In the graph-level contrast task, inter-sample learning is based on node-level and graph-level features. In the contrastive learning process, based on the defined node-level and graph-level positive and negative samples, the model strives to minimize the distance between positive samples while maximizing the distance between negative samples, thereby ensuring the consistency of encoded embeddings of each positive sample and effectively distinguishing the features of negative samples. Therefore, by analyzing the feature representations of the original graph through the contrastive learning process from multiple views, these features can be used for subsequent predictions of cell line-drug reactions.

### Node level contrastive learning

In the node-level contrastive learning task, for node  $v_i$  in any view  $v$ , only node  $u_i$  in the generated view  $u$  is its corresponding positive sample, and these two nodes are considered as the current positive sample pair, while other nodes in these two views are considered as negative samples. For each node-level positive sample pair, this paper considers that negative samples have two different sources: intra-graph negative samples and inter-graph negative samples.

In the node-level contrastive learning task, this paper defines a discriminator  $\theta(u_i, v_i) = s(g(u_i), g(v_i))$ , where  $s()$  represents cosine similarity, and  $g()$  represents a two-layer MLP nonlinear projection function, thus enhancing the discriminative power of the discriminator. For node-level positive sample pairs, the distance between positive sample pairs is defined as  $C_{pos} = e^{\theta(u_i, v_i)/\tau}$ , where an exponential function is used to amplify the similarity between positive samples, and  $\tau$  is the temperature parameter. Similarly, the distance between intra-graph positive and negative sample pairs is defined as  $C_{neg}^{intra} = e^{\theta(u_i, u_k)/\tau}$ , and the distance between inter-graph positive and negative sample pairs as  $C_{neg}^{inter} = e^{\theta(u_i, v_k)/\tau}$ .

Based on the above definition, for each node level sample pair  $(u_i, v_i)$ , the learning objectives of the node level samples are defined as:

$$h(u_i, v_i) = \log \frac{C_{pos}}{C_{pos} + \sum_{k=1}^N C_{neg}^{intra} + \sum_{k=1}^N C_{neg}^{inter}} \tag{24}$$

The loss of other positive sample graph structures is defined the same as the above loss. In order to ensure the maximization of the overall goal, this article defines the goal of average graph level positive sample pairs as:

$$Loss_{nod} = \frac{1}{2N} \sum_{i=1}^N (h(u_i, v_i) + h(v_i, u_i)) \tag{25}$$

Defining node-level positive and negative samples in contrastive learning tasks enables the model to maximize its learning of effective feature representations from node-level positive samples, thereby enhancing the extraction of effective features from the original graph.

## Graph level contrastive learning

In the graph-level contrastive learning task, this study contrasts the features of the original graph  $G_{sen}$  with its resistance graph  $G_{Res}$  and their generated graphs, using both node-level and graph-level features to enhance the model's generalization ability. Similar to the node-level contrastive learning task, in the graph-level task, for any node  $v_i$  in a given view, only the graph-level feature  $V$  of  $v_i$ 's view  $v$  is its positive sample, while the graph-level features of other graph structures serve as negative samples for node  $v_i$ . This paper posits that the completely opposite topological structure in the resistance graph  $G_{Res}$  compared to the original graph  $G_{sen}$  represents the antagonistic response between cell lines and drugs. Using two distinctly different topological structures for contrastive learning can enhance the model's discriminative ability.

In the graph-level contrastive learning task, another discriminator  $\varphi(v_i, V) = \sigma(v_i^T W_V)$  is defined, where  $\sigma()$  represents a sigmoid non-linear function,  $W_V$  is a learnable scoring matrix, and  $\varphi()$ , constructed by  $\sigma()$ , serves as a contrast discriminator to estimate the similarity between node-level embeddings and graph-level embeddings.

For the graph level, this study aims to maximize the divergence between node-level features and graph-level features to achieve the objective of contrastive learning, which is defined as:

$$f(v_i, u_i, V) = \sum_{i=1}^N \log(\varphi(v_i, V)) + \sum_{i=1}^N \log(1 - \varphi(u_i, V)) \quad (26)$$

The loss for the rest of the graph structures is defined similarly to the above. To ensure the maximization of the overall objective, the paper defines the objective for average graph-level sample pairs as:

$$Loss_{gra} = \frac{1}{2N} \sum_{i=1}^N (f(v_i, u_i, V) + f(u_i, v_i, U)) \quad (27)$$

where  $u_i$  is the corresponding node feature in the generated graph to node  $v_i$ , and  $U$  is the graph-level feature. Estimating node-level features and graph-level features among graph-level negative samples can enable the model to learn graph-level features to the maximum extent.

In summary, in each training round, cell line and drug data are input into a nonlinear subspace to merge linear and nonlinear features; and based on the heterogeneity of cell lines and drugs, a heterogeneous graph is constructed and input into the GraphMorpher module for graph enhancement, followed by a graph reasoning process using a graph encoder to obtain node representations of the graph; using the graph encoder for inference of node representations, the original graph features are finally optimized through contrastive learning for drug resistance prediction tasks.

For the final prediction of drug resistance tasks, the final embeddings of cell line nodes and drug nodes are utilized, and the probability of their sensitive response is predicted using a scoring function with an inner product.

## Acknowledgments

This work has been sponsored by multiple projects and supported by the computing and data resources of Shanxi Provincial People's Hospital, as well as the professional knowledge of its employees.

## Author Contributions

**Conceptualization:** Yunyun Dong, Xiufang Feng.

**Data curation:** Yunyun Dong, Xiufang Feng.

**Funding acquisition:** Yunyun Dong.

**Methodology:** Yuanrong Zhang.

**Software:** Yiming Zhao.

**Supervision:** Yunyun Dong, Yuhua Qian, Xiufang Feng.

**Validation:** Yiming Zhao.

**Visualization:** Ziting Yang.

**Writing – original draft:** Yunyun Dong.

## References

1. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW et al., "Systematic identification of genomic markers of drug sensitivity in cancer cells," *Nature*, 2012 vol. 483, no. 7391, pp. 570–575. <https://doi.org/10.1038/nature11005> PMID: 22460902
2. Sheng J, Li F, Wong STC, "Optimal drug prediction from personal genomics profiles," *IEEE J Biomed Health Inform*, 2015 vol. 19, no. 4, pp. 1264–1270. <https://doi.org/10.1109/JBHI.2015.2412522> PMID: 25781964
3. "Baptista D, Ferreira PG, Rocha M. Deep learning for drug response prediction in cancer. *Brief Bioinform*. 2021; 22(1):360–79. <https://doi.org/10.1093/bib/bbz171> WOS:000634950200028." PMID: 31950132
4. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S et al., "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Res*, 2012 vol. 41, no. D1, pp. D955–D961. <https://doi.org/10.1093/nar/gks1111> PMID: 23180760
5. Li Q, Shi R, Liang F, "Drug sensitivity prediction with high-dimensional mixture regression," *PLoS One*, 2019 vol. 14, no. 2, p. e0212108. <https://doi.org/10.1371/journal.pone.0212108> PMID: 30811440
6. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S et al., "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, 2012 vol. 483, no. 7391, pp. 603–607. <https://doi.org/10.1038/nature11003> PMID: 22460905
7. Wang L, Li X, Zhang L, Gao Q, "Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization," *BMC Cancer*, 2017 vol. 17, no. 1, pp. 1–12.
8. Geeleher P, Cox NJ, Huang RS, "Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines," *Genome Biol*, 2014 vol. 15, pp. 1–12. <https://doi.org/10.1186/gb-2014-15-3-r47> PMID: 24580837
9. Dong Z, Zhang N, Li C, Wang H, Fang Y, Wang J et al., "Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection," *BMC Cancer*, 2015 vol. 15, no. 1, pp. 1–12. <https://doi.org/10.1186/s12885-015-1492-6> PMID: 26121976
10. Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z et al., "Modeling precision treatment of breast cancer," *Genome Biol*, 2013 vol. 14, pp. 1–14. <https://doi.org/10.1186/gb-2013-14-10-r110> PMID: 24176112
11. Xia Z, Wu L-Y, Zhou X, Wong STC, "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces," in *BMC systems biology*, BioMed Central, 2010, pp. 1–16. <https://doi.org/10.1186/1752-0509-4-S2-S6> PMID: 20840733
12. LaFramboise T, "Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances," *Nucleic Acids Res*, 2009 vol. 37, no. 13, pp. 4181–4193. <https://doi.org/10.1093/nar/gkp552> PMID: 19570852
13. Bolton EE, Wang Y, Thiessen PA, Bryant SH, "PubChem: integrated platform of small molecules and biological activities," in *Annual reports in computational chemistry*, vol. 4, Elsevier, 2008, pp. 217–241.
14. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P, "Molecular graph convolutions: moving beyond fingerprints," *J Comput Aided Mol Des*, 2016 vol. 30, pp. 595–608. <https://doi.org/10.1007/s10822-016-9938-8> PMID: 27558503

15. Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester M, “MOLI: multi-omics late integration with deep neural networks for drug response prediction,” *Bioinformatics*, 2019 vol. 35, no. 14, pp. i501–i509. <https://doi.org/10.1093/bioinformatics/btz318> PMID: 31510700
16. Su R, Liu X, Wei L, Zou Q, “Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response,” *Methods*, 2019 vol. 166, pp. 91–102. <https://doi.org/10.1016/j.jymeth.2019.02.009> PMID: 30772464
17. Rahman R, Otridge J, Pal R, “IntegratedMRF: random forest-based framework for integrating prediction from different data types,” *Bioinformatics*, 2017 vol. 33, no. 9, pp. 1407–1410. <https://doi.org/10.1093/bioinformatics/btw765> PMID: 28334269
18. Ding MQ, Chen L, Cooper GF, Young JD, Lu X, “Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics,” *Molecular cancer research*, 2018 vol. 16, no. 2, pp. 269–278. <https://doi.org/10.1158/1541-7786.MCR-17-0378> PMID: 29133589
19. Zhang N, Wang H, Fang Y, Wang J, Zheng X, Liu XS, “Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model,” *PLoS Comput Biol*, 2015 vol. 11, no. 9, p. e1004498. <https://doi.org/10.1371/journal.pcbi.1004498> PMID: 26418249
20. Nguyen T, Nguyen GTT, Nguyen T, Le D-H, “Graph convolutional networks for drug response prediction,” *IEEE/ACM Trans Comput Biol Bioinform*, 2021 vol. 19, no. 1, pp. 146–154.
21. Liu Q, Hu Z, Jiang R, Zhou M, “DeepCDR: a hybrid graph convolutional network for predicting cancer drug response,” *Bioinformatics*, 2020 vol. 36, no. Supplement\_2, pp. i911–i918.
22. Li M, Wang Y, Zheng R, Shi X, Li Y, Wu F-X et al., “DeepDSC: a deep learning method to predict drug sensitivity of cancer cell lines,” *IEEE/ACM Trans Comput Biol Bioinform*, 2019 vol. 18, no. 2, pp. 575–582.
23. Choi J, Park S, Ahn J, “RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance,” *Sci Rep*, 2020 vol. 10, no. 1, p. 1861. <https://doi.org/10.1038/s41598-020-58821-x> PMID: 32024872
24. Liu P, Li H, Li S, Leung K-S., “Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network,” *BMC Bioinformatics*, 2019 vol. 20, no. 1, pp. 1–14.
25. Li Q, Huang J, Zhu H, Liu Q, “Prediction of cancer drug effectiveness based on multi-fusion deep learning model,” in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2020, pp. 634–639.
26. Turki T, Wei Z, “A link prediction approach to cancer drug sensitivity prediction,” *BMC Syst Biol*, 2017 vol. 11, pp. 1–14.
27. Zhang F, Wang M, Xi J, Yang J, Li A, “A novel heterogeneous network-based method for drug response prediction in cancer cell lines,” *Sci Rep*, 2018 vol. 8, no. 1, p. 3355. <https://doi.org/10.1038/s41598-018-21622-4> PMID: 29463808
28. Zhu Y, Xu Y, Yu F, Liu Q, Wu S, Wang L, “Graph contrastive learning with adaptive augmentation,” in *Proceedings of the Web Conference 2021*, 2021, pp. 2069–2080.
29. Xu D, Cheng W, Luo D, Chen H, Zhang X, “Infogcl: Information-aware graph contrastive learning,” *Adv Neural Inf Process Syst*, 2021 vol. 34, pp. 30414–30425.
30. Chen T, Kornblith S, Norouzi M, Hinton G, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*, PMLR, 2020, pp. 1597–1607.
31. Veličković P, Fedus W, Hamilton WL, Liò P, Bengio Y, Hjelm RD, “Deep graph infomax,” *arXiv preprint arXiv:1809.10341*, 2018.
32. Qiu J, Chen Q, Dong Y, Zhang J, Yang H, Ding M et al., “Gcc: Graph contrastive coding for graph neural network pre-training,” in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 1150–1160.
33. Liu X, Song C, Huang F, Fu H, Xiao W, Zhang W, “GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction,” *Brief Bioinform*, 2022 vol. 23, no. 1, p. bbab457. <https://doi.org/10.1093/bib/bbab457> PMID: 34727569
34. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W et al., “A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information,” *Nat Commun*, 2017 vol. 8, no. 1, p. 573. <https://doi.org/10.1038/s41467-017-00680-8> PMID: 28924171
35. Fu H, Huang F, Liu X, Qiu Y, Zhang W, “MVGCN: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks,” *Bioinformatics*, 2022 vol. 38, no. 2, pp. 426–434. <https://doi.org/10.1093/bioinformatics/btab651> PMID: 34499148
36. Huang S, Hu P, Lakowski TM, “Predicting breast cancer drug response using a multiple-layer cell line drug response network model,” *BMC Cancer*, 2021 vol. 21, no. 1, p. 648. <https://doi.org/10.1186/s12885-021-08359-6> PMID: 34059012

37. Ammad-Ud-Din M, Khan SA, Malani D, Murumägi A, Kallioniemi O, Aittokallio T et al., “Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization,” *Bioinformatics*, 2016 vol. 32, no. 17, pp. i455–i463. <https://doi.org/10.1093/bioinformatics/btw433> PMID: 27587662
38. Suphavilai C, Bertrand D, Nagarajan N, “Predicting cancer drug response using a recommender system,” *Bioinformatics*, 2018 vol. 34, no. 22, pp. 3907–3914. <https://doi.org/10.1093/bioinformatics/bty452> PMID: 29868820
39. Kipf TN, Welling M, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.
40. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, “Graph attention networks,” arXiv preprint arXiv:1710.10903, 2017.
41. Peng W, Liu H, Dai W, Yu N, Wang J, “Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions,” *Bioinformatics*, 2022 vol. 38, no. 19, pp. 4546–4553. <https://doi.org/10.1093/bioinformatics/btac574> PMID: 35997568
42. Peng W, Chen T, Dai W, “Predicting drug response based on multi-omics fusion and graph convolution,” *IEEE J Biomed Health Inform*, 2021 vol. 26, no. 3, pp. 1384–1393.
43. Yao K, Wang X, Li W, Zhu H, Jiang Y, Li Y et al., “Semi-supervised heterogeneous graph contrastive learning for drug–target interaction prediction,” *Comput Biol Med*, 2023 vol. 163, p. 107199. <https://doi.org/10.1016/j.compbiomed.2023.107199> PMID: 37421738
44. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J et al., “Self-supervised learning: Generative or contrastive,” *IEEE Trans Knowl Data Eng*, 2021 vol. 35, no. 1, pp. 857–876.
45. Li Y, Qiao G, Gao X, Wang G, “Supervised graph co-contrastive learning for drug–target interaction prediction,” *Bioinformatics*, 2022 vol. 38, no. 10, pp. 2847–2854. <https://doi.org/10.1093/bioinformatics/btac164> PMID: 35561181
46. Wang Y, Min Y, Chen X, Wu J, “Multi-view graph contrastive representation learning for drug-drug interaction prediction,” in *Proceedings of the Web Conference 2021*, 2021, pp. 2921–2933.
47. Zhuang L, Wang H, Zhao J, Sun Y, “Adaptive dual graph contrastive learning based on heterogeneous signed network for predicting adverse drug reaction,” *Inf Sci (N Y)*, 2023 vol. 642, p. 119139.
48. Xu R, Luo L, Liu Z, Ouyang C, Wan Y, “CLDDI: A Novel Method for Predicting Drug-Drug Interaction Events Based on Graph Contrastive Learning,” in *2023 11th International Conference on Bioinformatics and Computational Biology (ICBCB)*, IEEE, 2023, pp. 105–112.
49. Wei J, Zhuo L, Zhou Z, Lian X, Fu X, Yao X, “GCFMCL: predicting miRNA-drug sensitivity using graph collaborative filtering and multi-view contrastive learning,” *Brief Bioinform*, 2023 vol. 24, no. 4, p. bbad247. <https://doi.org/10.1093/bib/bbad247> PMID: 37427977
50. Zhou H, Gong M, Wang S, Gao Y, Zhao Z, “SMGCL: Semi-supervised Multi-view Graph Contrastive Learning,” *Knowl Based Syst*, 2023 vol. 260, p. 110120.
51. Goldschmidt H, Lokhorst HM, Mai EK, van der Holt B, Blau IW, Zweegman S et al., “Bortezomib before and after high-dose therapy in myeloma: long-term results from the phase III HOVON-65/GMMG-HD4 trial,” *Leukemia*, 2018 vol. 32, no. 2, pp. 383–390, <https://doi.org/10.1038/leu.2017.211> PMID: 28761118
52. You Y, Chen T, Sui Y, Chen T, Wang Z, Shen Y, “Graph contrastive learning with augmentations,” *Adv Neural Inf Process Syst*, 2020 vol. 33, pp. 5812–5823.
53. Jin W, Zhao L, Zhang S, Liu Y, Tang J, Shah N, “Graph condensation for graph neural networks,” arXiv preprint arXiv:2110.07580, 2021.
54. Bonacich P, “Power and centrality: A family of measures,” *American journal of sociology*, 1987 vol. 92, no. 5, pp. 1170–1182.
55. Brin S, “The PageRank citation ranking: bringing order to the web,” *Proceedings of ASIS*, 1998, 1998 vol. 98, pp. 161–172.