OXFORD

Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism

Peng Zhang [ID], Shikui Tu, Wen Zhang and Lei Xu

Corresponding authors: Shikui Tu, Department of Computer Science and Engineering, Center for Cognitive Machines and Computational Health (CMaCH), Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: tushikui@sjtu.edu.cn; Wen Zhang, Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, College of informatics, Huazhong Agricultural University, Wuhan 430070, China. E-mail: zhangwen@whu.edu.cn

Abstract

Identifying synergistic drug combinations (SDCs) is a great challenge due to the combinatorial complexity and the fact that SDC is cell line specific. The existing computational methods either did not consider the cell line specificity of SDC, or did not perform well by building model for each cell line independently. In this paper, we present a novel encoder-decoder network named SDCNet for predicting cell line-specific SDCs. SDCNet learns common patterns across different cell lines as well as cell line-specific features in one model for drug combinations. This is realized by considering the SDC graphs of different cell lines as a relational graph, and constructing a relational graph convolutional network (R-GCN) as the encoder to learn and fuse the deep representations of drugs for different cell lines. An attention mechanism is devised to integrate the drug features from different layers of the R-GCN according to their relative importance so that representation learning is further enhanced. The common patterns are exploited through partial parameter sharing in cell line-specific decoders, which not only reconstruct the known SDCs but also predict new ones for each cell line. Experiments on various datasets demonstrate that SDCNet is superior to state-of-the-art methods and is also robust when generalized to new cell lines that are different from the training ones. Finally, the case study again confirms the effectiveness of our method in predicting novel reliable cell line-specific SDCs.

Keywords: drugs, synergistic drug combinations, attention mechanism, relational graph convolutional network, cancer treatment

Introduction

Limited effects are achieved for the monotherapies with few exceptions for complex diseases [1]. Drug combination therapy, which uses multiple drugs concurrently or sequentially, is a widely used paradigm for complex diseases such as cancers [2, 3]. However, not all drug combinations are synergistic drug combinations (SDCs) that can increase therapeutic efficacy and overcome drug resistance [4, 5]. Some combinations may cause severe side effects [6, 7]. Therefore, it is a crucial task to determine the synergistic combinations that have greater effects than single-drug administration [8, 9]. This task is difficult due to the vast number of possible combinations and the fact that drug combination therapy has cell line-specific responses [10, 11].

Traditional methods for discovering novel SDCs are empirical exercises based on available biological relationships for oncology pathway or clinical trials [12, 13]. These methods are time-consuming and experimentally expensive because of the numerous combinations of candidates [14, 15]. With the rapid development of high-throughput screening (HTS) techniques, it becomes possible to experimentally evaluate the synergistic effects of abundant drug combinations across different cell lines, facilitating the discovery of rational combination therapies for patients [16, 17]. For example, an unbiased screen of 22 737 sensitivity experiments has been undertaken for 583 doublet drug combinations in 39 diverse cancer cell lines, which identified many novel synergistic and efficacious drug combinations as potential candidates to aid in drug development [18]. Computational methods are also valuable tools to prioritize the SDCs and predict high-confidence candidates for experimental validation. The searching range of potential combinations is narrowed down, and thus it saves time and cost for real biological experiments. The existing drug combination databases such

Peng Zhang is a postdoc fellow in the Department of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU). His research interests include machine learning and bioinformatics.

Shikui Tu, now a tenure-track associate professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU). His research interests include machine learning and bioinformatics. He has published more than 40 academic papers in top conferences and journals, including Science, Cell, NAR, etc.

Wen Zhang obtained a bachelor's degree and a master's degree in computational mathematics from Wuhan University in 2003, 2006 and got a doctoral degree in computer science from Wuhan University in 2009. He is now a professor in the College of Informatics, Huazhong Agricultural University, People's Republic of China. His research interests include bioinformatics and machine learning.

Lei Xu, is a Zhiyuan Chair Professor of Computer Science and Engineering Department, chief scientist of AI Research Institute, chief scientist of Brain Sci & Tech Research Centre, Shanghai Jiao Tong University (SJTU) and director of Neural Computation Research Centre in Brain and Intelligence Science-Technology Institute, Zhang Jiang National Lab. Elected to Fellow of IEEE in 2001 and Fellow of IAPR 2002 and of European Academy of Sciences (EURASC) in 2003.

as DCDB [19], DrugComb [20] and DrugCombDB [21] provide unprecedented opportunities to build reliable computational methods for identifying novel and efficient SDCs for the patients [22, 23].

In recent years, a number of computational methods have been proposed to predict novel SDCs [24, 25]. DeepSynergy [26] and Matchmaker [27] adopt fully connected layers to construct neural network models for SDC prediction, where drug chemical features and gene expression profiles from every cell line are taken as input. Different from the chemical information-based drug representation, TransSynergy [28] employs a transformer network and considers target-based vector representations as input, which are computed from drug-target and protein–protein interaction profiles via random walk algorithms. The target-based drug representations are more closely related to the cellular response to the drug treatment, and thus improve the model performance. Although all these methods take cell-line specificity into account, they rely on extra data such as gene expression or loss-of-function screening profiles. One way of removing the dependence on the extra cell line data is to represent the drug combinations' synergy data as a multidimensional array or tensor. DTF combines a tensor factorization method and neural network, successfully capturing the structure of multidimensional tensor, which is constructed from the drug combination's synergy data, to predict the synergistic effects of drug combinations [29]. Although the above methods have achieved incredible performances for predicting novel SDCs, they still have room for improvement.

Recently, graph neural network (GNN) and graph convolutional network (GCN) become popular in the bioinformatics field [30]. GNN-based models have been built for various drug-related tasks, with the advantage of capturing the dependence within the graph-structured drug data through message propagation [31]. Jiang et al. trained GCN encoders for every cell line on the heterogonous network data that are constructed from drug–drug synergy subnetwork, drug-protein interactions and protein–protein interactions, and used cell line-specific matrix bilinear decoder to predict the drug–drug synergy [10]. Wang et al. treated the drug chemical data as graphs where the vertices are atoms and the edges are chemical bonds, designed a GCN-based network to learn the deep representations of drugs, and deployed multi-layer perception (MLP) to extract the cell line features from gene expression profiles, altogether forming a pipeline called DeepDDS to predict novel SDCs [32]. However, these methods can effectively exploit the drug combination's unique features in specific cell lines, but do not consider the invariant patterns among the cell lines. In particular, the drug combinations in different cell lines are highly associated with each other [18, 33]. Hence, the invariant patterns of drug combinations among cell lines cannot be ignored, and models merging these common features will benefit their prediction of cell line-specific SDCs.

To solve the above drawbacks and limitations, in this paper, we present a novel GCN-based method named SDCNet to predict cell line-specific SDCs, without resorting to any cell line data such as gene expressions. SDCNet is able to learn the common features of drug combinations across cell lines improving the prediction accuracy for each cell line. We represent the synergy data as graphs for every cell line, and integrate the graphs to obtain a relational graph, where cell lines are regarded as different types of relations. We model the relational graph by relational graph convolutional network (R-GCN) [34], which can simultaneously capture the nodes' unique features on each single relation and their invariant patterns across different relations. Although R-GCN has been used in many biomedical network analyses such as

multicategory microRNA (miRNA)–disease association prediction [35], multiple types of drug-target interaction prediction [36] and polypharmacy side-effect prediction [37], it has not been investigated on the problem of SDC prediction. Moreover, we present a layer attention mechanism to fuse the features of different levels of network layers. Experiments on various datasets demonstrate that SDCNet is superior to the state-of-the-art methods. It is also found that the drug embedding can be transferred via SDCNet from one dataset to another to enhance the generalization performance.

Materials
Datasets

The drug combinations' synergy data are mainly from four datasets, namely O'Neil, ALMANAC, CLOUD and FORCINA, in the DrugComb database [20, 38]. From the O'Neil dataset, the data from 31 cell lines are used to compare the performances of SDCNet and existing methods on SDC prediction [18]. The data of the remaining eight cell lines are assembled as the transfer dataset to evaluate the performance of transfer learning on SDC prediction. For the other three datasets, all data are used to evaluate the generalization performances of the methods except some samples are excluded due to lack of the molecular information of drugs. The synergistic effects of the drug combinations are quantified through four types of synergy scores including Loewe additivity (Loewe) [39], bliss independence (Bliss) [40], zero interaction potency (ZIP) [41] and highest single agent (HSA) [42]. Notably, the experimental replicates of the drug combinations in the O'Neil dataset are averaged separately for various synergy types. According to the synergy scores, a drug combination can be roughly classified into synergistic drug combination and antagonistic drug combination with a threshold of zero [21]. Drug combinations who with higher synergy scores are more synergistic, and vice versa. However, the synergy scores of most drug combinations are approximately zero due to the noise in the HTS experiments (additive drug combinations) and they are hard to classify [21]. To address this problem, researchers tend to select a relatively strict threshold to exclude the low-confidence drug combinations [10, 26]. The four synergy types, which are identified under different empirical or biological assumptions, quantify the degree of interaction in different ways. Therefore, the thresholds for classifying the synergistic drug combination (positive) and antagonistic drug combination (negative) are different from each other. In particular, the drug combinations with Loewe scores higher than 10 are treated as positive samples, those lower than 0 are treated as negative samples, while the other samples are excluded following the previous study [32]. To evaluate the impact of different synergy thresholds on Loewe scores, we also conduct experiments on the synergy data using the threshold of 30 to label the combination as synergistic, the same as the setting in [26]. The data and results are given in Supplementary Table S1. For the other three synergy types, the quartiles of the samples in the DrugComDB database are adopted as thresholds [21]. Specifically, the thresholds are {−3.37, 3.68} for the Bliss score, {−3.02, 3.87} for the HSA score, and {−4.48, 2.64} for the ZIP score, where the samples with scores higher than the large value are positive samples, the samples with scores lower than the small value are negative samples, and the other samples are excluded. Table 1 summarizes the drug combinations' synergy data in different datasets with various synergy types, and Supplementary Tables S2-S4 summarize the data in each cell line for different datasets.

Table 1. Summary of the drug combinations' synergy data for different datasets with various synergy types

Synergy type	Samples	O'Neil dataset	ALMANAC dataset	CLOUD dataset	FORCINA dataset	Transfer dataset
	Drugs	38	82	242	757	38
	Cell lines	31	59	1	1	8
	All combinations	18,071	154,596	29,278	757	4664
Loewe	Positive	6188	1950	5365	273	1634
	Negative	7055	122,574	17,160	165	1622
Bliss	Positive	6166	30,973	5308	477	1663
	Negative	3751	29,541	20,133	93	675
ZIP	Positive	7007	29,539	6058	504	1799
	Negative	2618	34,330	19,720	115	516
HSA	Positive	9437	21,514	9078	479	2521
	Negative	1897	41,161	14,989	128	322

Drug features

The simplified molecular input line entry system (SMILES) of drugs were downloaded from DrugBank [43]. Then, the drug feature is represented by the molecular fingerprint, which is a numerical vector of specific length calculated based on its SMILES [44]. In our study, a 300 bits long Infomax fingerprint is used as the drug feature. Zagidullin et al. has systematically evaluated the performance of 3 rule-based (Topological, Morgan and E3FP) and 4 DL-based (GAE, VAE, transformer and Infomax) molecular fingerprints for the prediction of drug synergy scores and found that the best performance is achieved by Infomax fingerprints compared with the other representations [45]. Specifically, the Infomax fingerprints are generated using a pre-trained Deep Graph Infomax model, which has been pre-trained on millions of molecules from ChEMBL 20 and ZINC 15 by Hu et al. [21, 46]. Finally, the drugs features are integrated into the SDC graph of each cell line, providing diverse molecular information about the drugs.

Construction of the SDC graphs and relational graph

The SDC graph $G_r = (V, \mathcal{E})$ of cell line $r \in R$ is first constructed by using the drug combinations' synergy data, where V is the set of nodes (drugs), $\mathcal{E}$ is the set of labeled edges (known SDCs in cell line r) and R represents the set of cell lines (Figure 1A). The adjacency matrix of the SDC graph G_r is defined as $A_r \in [0, 1]^{N \times N}$, where N denotes the number of drugs. $A_r(i, j)$ is set to 1 if the combination between drug i and drug j in cell line r is synergistic; otherwise, $A_r(i, j)$ is 0. Since the SDC graph is undirected, $A_r(i, j) = A_r(j, i)$. Then, SDC graphs of all cell lines are merged into a relational graph $G = (V, \mathcal{E}, R)$ (Figure 1A), a type of graph in which edges have multiple labels. The relational graph is a more general and pervasive class of graph compared with simple homogeneous graph [34, 47]. In this paper, different cell lines of the drug combinations' synergy data are regarded as multiple categories of relations in the relational graph. More precisely, the label of the combination between drug i and drug j in relation r is 1 if it is SDC; otherwise, it is 0.

Method

We formulate the prediction of cell line-specific SDCs as a link prediction problem. An overview of the proposed SDCNet is given in Figure 1. SDCNet is an encoder-decoder network. The drug combinations' synergy data are represented as SDC graphs, with nodes representing drugs and edges representing SDCs. The SDC graphs of different cell lines are merged into a relational graph by regarding the cell lines as various relations. For the graph-structured data, GCN is an effective method due to its advantage of capturing the dependence and learning low-dimensional representations for the nodes through message propagation [48]. For the relational graph, R-GCN, a variant of GCN, is more efficient because it can simultaneously capture the nodes' unique features on each single relation and their invariant patterns across different relations [34]. Hence, the encoder adopts a R-GCN to learn the embeddings of synergy patterns across the cell lines for every drug, and then the decoder reconstructs the known cell line-specific SDCs and predicts new ones from the drug embedding. Therefore, the synergy data from all cell lines are jointly taken into account in one model, which captures not only cell line-specific information but also cell line-invariant synergy features and requires no extra cell line genomic data.
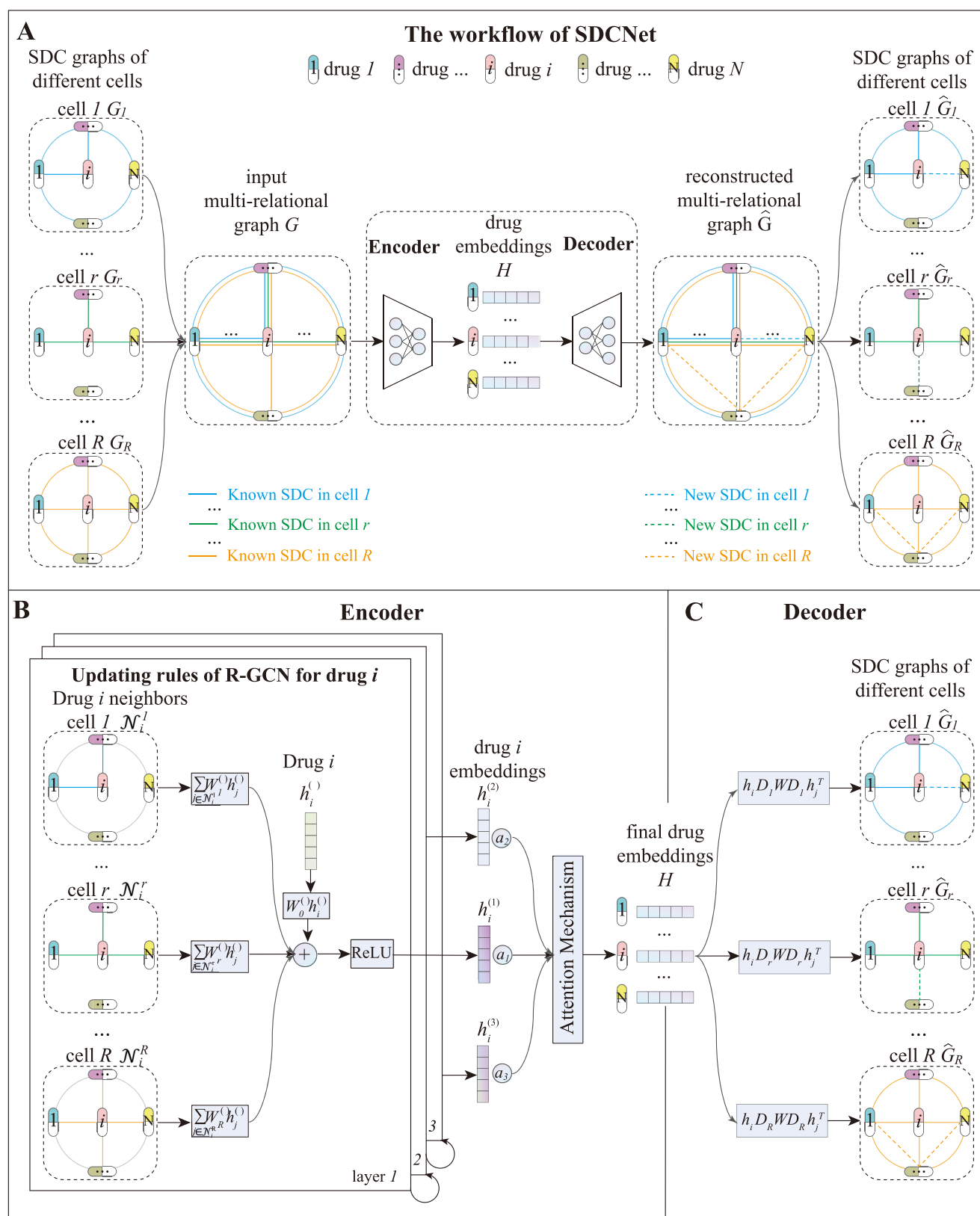
Encoder

The encoder stacks multiple R-GCN layers to learn the low-dimensional drug embedding by fusing the drug combinations' unique features in a specific cell line and their common features across different cell lines (Figure 1B). Specifically, the update rule of R-GCN for drug i is defined by Equation (1):

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right), \qquad (1)$$

where $h_i^{(l)} \in \mathbb{R}^{1 \times d}$ is the embedding of drug i at the l-th layer and d means the dimensionality of the drug embedding. $\mathcal{N}_i^r$ denotes the set of drug i neighbors in SDC graph G_r. $c_{i,r}$ is a normalization constant that can either be learned or chosen in advance (e.g. $|\mathcal{N}_i^r|$) to weigh the contribution of drug i neighbors from cell line r. $|\mathcal{N}_i^r|$ is the number of drug i neighbors in SDC graph G_r. $W_r^{(l)}$ represents a trainable cell line-specific weight matrix at the l-th layer, while $W_0^{(l)}$ is a weight matrix for drug i itself at the l-th layer. And σ is the ReLU activation function. With the above settings, we initialize the drug embedding $h_i^{(0)}$ with the drug molecular feature.

For the relational graph data, stacking multiple R-GCN layers can make the model be more expressive and aware of the graph structure [49]. Each layer learns a certain level of structural information in the graph network for the nodes. Specifically, the l-th layer can encodes the l-hop neighbor information of the nodes [30]. Therefore, the drug embeddings from each layer may contribute in different levels to the final prediction. We employ

Figure 1. The overview of SDCNet. (A) The SDC graphs of different cell lines, the relational graph and the workflow of SDCNet. (B) The schematic of encoder in SDCNet. (C) The schematic of decoder in SDCNet. N is the number of drugs and R is the set of cell lines.

the attention mechanism to place different weights for each layer's embedding. The final drug embedding h_i is calculated using Equation (2):

$$h_i = \sum_{l \in L} \alpha_l \times h_i^{(l)}, \tag{2}$$

where α_l is a trainable constant that indicates the attention weight of the l-th R-GCN layer. $l \in L$, L is the set of R-GCN layers.[[ineq32]]We use $H \in \mathbb{R}^{N \times d}$ to represent the drug embeddings matrix, which stacks all the drugs embeddings.

Decoder

Taking the final drug embedding learned by the encoder as input, a decoder is adopted to reconstruct the SDC graph of each cell line (Figure 1C). In particular, the decoder assigns a probability score p to a sample, which represents how likely the combination is SDC in the corresponding cell line. More precisely, utilizing the embedding vectors of drug i and drug j learned by the encoder, the decoder scores their combination in the cell line r (i, j, r) through a factorized operation:

$$p_r^{ij} = p(i, j, r) = \sigma\left(h_i D_r W D_r h_j^T\right) \tag{3}$$

where $W \in \mathbb{R}^{d \times d}$ is a trainable weight matrix, which is shared in different cell lines, modeling global interactions of drug combinations across different cell lines. $D_r \in \mathbb{R}^{d \times d}$ is a diagonal matrix that captures the importance of each dimension in drug embeddings h_i and h_j towards the cell line r, and σ is the sigmoid function. The decoder is in a matrix bilinear form, so it is also known as bilinear decoder [35, 37].

Training the model

Our SDCNet method learns the low-dimensional drug embedding in the encoder, and feeds them into the decoder to simultaneously predict the potential SDCs in different cell lines. This means that the loss of SDCNet merges all of the losses in each cell line. However, the numbers of positive samples and negative samples vary in different cell lines, and some are even extremely unbalanced, e.g. the data in the ALMANAC dataset with Loewe scores (Supplementary Table S3). Hence, we adopt the weighted cross-entropy as the loss function in each cell line, and the overall loss for SDCNet is defined by Equation (4):

$$Loss = -\sum_{r \in R} \frac{1}{M_r} \left(\lambda_r \times \sum_{(i,j) \in y_r^+} \log p_r^{ij} + \sum_{(i,j) \in y_r^-} \log\left(1 - p_r^{ij}\right) \right), \tag{4}$$

where (i, j) denotes the combination of drug i and drug j in the cell line r, and y_r^+ and y_r^- represent the set of positive and negative samples in the cell line r, respectively. The balance factor $\lambda_r = \frac{|y^-|}{|y^+|}$ imposes the importance of positive samples to reduce the impact of data imbalance, where $| y_r^+ |$ and $| y_r^- |$ are the number of combinations in y_r^+ and y_r^-, respectively. M_r is the number of samples in the cell line r required to normalize the loss value from the cell line r.

We deploy an end-to-end optimization approach, which has been shown to significantly improve model performance on graph-structured data [35, 50]. All of the trainable parameters involved in the encoder and decoder are first initialized by the Xavier uniform initialization method [51], and jointly optimized via a gradient descent with the Adam optimizer [52]. Additionally, we add a regular dropout in the graph convolution layer unit, helping the model generalized well to unknown edges [53]. Finally,

to balance the training speed and accuracy during optimization, we apply a cyclic learning rate, which makes a change between the maximum and minimum learning rate [54]. The models are implemented by TensorFlow (version 2).

Results and discussion
Experimental setting

To evaluate the performances of different methods, we adopt 10-fold cross-validation. In each cell line, the positive samples are randomly split into ten equal-sized subsets; one subset is selected and concatenated with the same number of negative samples as the test data, while the remaining positive and negative samples are taken as training data. Then, the training data and test data from all cell lines are merged as the final training dataset and test dataset, respectively. The training dataset is used to construct the relational graph and train the model, and the test dataset is used to predict the potential SDC in specific cell lines. We iteratively repeat the cross-validation process ten times to make full use of the datasets. In addition, the commonly used evaluation metrics, including the area under the curve (AUC), accuracy (ACC), area under the precision recall (AUPR), F1 score (F1), precision and recall are separately calculated for each cell line to measure the performance of a model in a specific cell line. The metric values in all cell lines are separately averaged for each evaluation metric to measure the performance of the model for all cell lines. For each evaluation metric, we report the mean value and standard deviation of the results on each fold of the 10-fold cross-validation data.

There are several hyperparameters in SDCNet such as the layers of the R-GCN l, dimensionality of the drug embeddings d, the number of training epochs, learning rate, and dropout rate. We consider different combinations of these parameters: l in the range of {1, 2, 3, 4}; d in the range of {64, 128, 192, 256, 320, 384}; the number of training epochs in the range of {6000, 8000, 10 000, 12 000, 14 000}; the learning rate in the range of {10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}}; and the dropout rate in the range of {0.1, 0.2, 0.3, 0.4, 0.5}. After tunning the hyperparameters via 10-fold cross-validations on the O'Neil dataset with Loewe scores, we set the parameters as $l = 3$, $d = 320$, the number of training epochs = 10 000, learning rate = 10^{-3} and dropout = 0.2.

Comparison with state-of-the-art methods

To evaluate the performance of SDCNet, we compare it with several state-of-the-art methods on various datasets with four synergy types (Loewe, Bliss, HSA and ZIP), respectively. Furthermore, the hyperparameters of these methods are allowed to be adjusted with grid search via 10-fold cross-validation, and the optimal values are selected as their final performance.

- DeepSynergy [26] constructed an NN model that combined both drug chemical features and cell line genomic features to predict the synergistic scores of drug combinations.
- Matchmaker [27] further learned the drug-specific features by NN compared with DeepSynergy.
- DTF [29] first extracted latent representations of drugs and cell lines from the multidimensional tensor, which is constructed from the drug combination's synergy data, by tensor factorization, then predict the synergistic effects of drug combinations through NN.
- DeepDDS [32] treated the drug chemical substructure as a graph to learn the corresponding drug representation by

GNN and integrate the cell line genomic features predicting novel SDCs.
- Jiang's method [10] leveraged heterogonous graph embeddings to learn drug representations and predict novel cell line-specific SDCs.

Table 2 summarizes the performances of different methods for all cell lines on the O'Neil dataset with four synergy types. For the data determined by the Loewe score, the best AUC, ACC AUPR and F1 values of 0.93, 0.85, 0.92 and 0.83, are achieved using SDCNet, respectively. Second best performance is achieved by the tensor factorization-based method, DTF, demonstrating its huge potential in capturing the latent representations for different variables. Similar results are obtained using GNN-based methods, including DeepDDS and Jiang's method, indicating that GNN can learn useful topological information from the graph-structured data. When comparing all methods' results on datasets with various synergy types, SDCNet always obtain the best performance for all evaluation metrics indicating that SDCNet generalizes across different synergy types well. Furthermore, we tested the performances of different methods on the other three datasets, respectively. The results show that SDCNet outperforms other methods on the ALMANAC dataset across all synergy types in terms of ACC, AUPR or F1 value (Table 3). Considering the results on the CLOUD and FORCINA datasets in Supplementary Tables S5 and S6, we have two observations. Firstly, SDCNet is a competing method on the single-cell line datasets. All methods show their relative strengths and weaknesses and no method is consistently superior to the others. For instance, on the CLOUD dataset with Loewe score, SDCNet achieves the highest F1 score of 0.37, while DeepDDS obtains the highest AUC and AUPR with values 0.67 and 0.67, respectively. Secondly, when comparing the results of SDCNet on different datasets, there is a drop in performance on single-cell line datasets, CLOUD and FORCINA. Taking the Loewe score data as example, SDCNet achieves the best F1 scores of 0.83 and 0.67 on the O'Neil and ALMANAC datasets, respectively, while the F1 scores drop to 0.37 and 0.55 on the CLOUD and FORCINA datasets. The strengths of SDCNet come from learning common patterns across different cell lines as well as cell line-specific features in one model for drug combinations. When training and testing on a single cell line dataset, FORCINA or CLOUD, there is no common drug embedding across different cell lines to be exploited in the decoder. Thus, SDCNet losses its strength on single-cell line datasets when comparing with the existing methods. This is consistent with the results in Figure 3. In Figure 3, when one model per cell line is trained and tested, it leads to consistent performance drops for every cell line in comparisons with one model for all cell lines. We also include an experiment of training and testing the models on the merged dataset of all the four benchmark datasets. Supplementary Table S7 summaries the results of different methods on the merged dataset with Loewe score. The results again confirm that SDCNet outperforms other methods in terms of most evaluation metrics with the highest AUC of 0.79, ACC of 0.68, AUPR of 0.80 and F1 value of 0.70, respectively. Additionally, it is worth noting that unlike the other methods, SDCNet does not use the corresponding cell line genomic features, which further indicates the effectiveness of SDCNet. To summarize, the results of methods on different datasets with four synergy types demonstrate the robustness and generality of SDCNet on datasets from multi-cell lines.

Moreover, to evaluate the predictive and generalization performance on novel drugs or cell lines, we test the performance of SDCNet on the O'Neil dataset (Loewe) with two more different strategies including i.e. leave one drug out, and leave one cell line out. The same strategies have been adopted by Preuer et al. [26] and Lin et al. [55]. The performances of various methods with different strategies are summarized in Supplementary Table S8. All methods achieve relatively low predictive performance when generalizing to novel drugs or cell lines, which are consistent with the results in Preuer et al. [26] and Lin et al. [55]. For the leave one drug out strategy, SDCNet performs well in terms of ACC, AUPR and precision, while DeepDDS achieves better performance with AUC and F1 values of 0.73 and 0.56, respectively. For the strategy of leave one cell line out, SDCNet obtains the best performance with an AUC, ACC, AUPR and F1 of 0.91, 0.84, 0.90 and 0.79, respectively. The results indicate the robustness and good generalization performance of SDCNet for SDC prediction in novel cell lines.

To further evaluate the performance of SDCNet in specific cell lines, we compare the metric values of different methods in each cell line on the O'Neil dataset (Loewe). The distributions of AUC and AUPR for different methods in each cell line are shown in Figure 2. Significantly, SDCNet achieves consistent and robust performance in each cell line, and its minimum AUC and AUPR values in the cell lines are 0.83 and 0.84 respectively. Among the 31 cell lines, SDCNet achieves the highest AUC in 15 cell lines and the highest AUPR in 12 cell lines, which is better than the other five methods. In addition, the performances of all methods varied in different cell lines (Supplementary Figure S1), which may be partially explained by the different number of drug combinations and rate between positive and negative samples in each cell line (Table 1).

Common features of drug combinations across cell lines improve SDC prediction

SDCNet captures not only cell line-specific information but also common features across cell lines for drug combinations. To evaluate whether the common features of drug combinations across cell lines can improve SDC prediction, we systematically compare the performances of SDCNet models with and without the common features in each cell line. Specifically, SDCNet takes into account the synergy data from all cell lines in one model, which means one model for all cell lines. In contrast, SDCNet without common features is one model for per cell line, which indicates that they can leverage only the drug combinations' unique information in a specific cell line.

The results of the SDCNet models with and without the common features on the O'Neil dataset determined by Loewe score (31 cell lines) are displayed in Figure 3. The difference between the two types of models is separately calculated for each metric, and the significance is determined by the Wilcoxon signed rank-sum test in every cell line. For the AUC and AUPR values, the SDCNet model obtained at least 1% improvement compared with the SDCNet models without common features in 28 and 29 cell lines, respectively, and the corresponding P-values are below 0.05. Similarly, the other evaluation metrics including ACC, F1 score, and recall also support the same trend, where SDCNet achieves superior performance in most cell lines (Supplementary Figure S2), and the detailed values are listed in Supplementary Tables S9 and S10. Based on the above results, we conclude that the common features of drug combinations among cell lines can significantly improve cell line-specific SDC prediction in most cell lines.

## Effect of attention mechanism in SDCNet

To quantify the effect of the attention mechanism in SDCNet, we compare the performance of SDCNet with five of its variants

Table 2. The performance of SDCNet and state-of-the-art methods on the O'Neil dataset with various synergy types

Synergy type	Methods	AUC	ACC	AUPR	F1	Precision	Recall
	DeepSynergy	0.80 ± 0.01	0.73 ± 0.01	0.78 ± 0.01	0.71 ± 0.01	0.71 ± 0.01	0.70 ± 0.02
	Matchmaker	0.81 ± 0.01	0.73 ± 0.01	0.79 ± 0.01	0.71 ± 0.01	0.70 ± 0.01	0.73 ± 0.02
Loewe	Jiang's method	0.86 ± 0.01	0.78 ± 0.01	0.86 ± 0.01	0.76 ± 0.01	0.80 ± 0.01	0.71 ± 0.02
	DeepDDS	0.89 ± 0.01	0.83 ± 0.01	0.86 ± 0.01	0.81 ± 0.01	0.81 ± 0.02	0.81 ± 0.01
	DTF	0.91 ± 0.01	0.82 ± 0.01	0.90 ± 0.01	0.81 ± 0.01	0.83 ± 0.03	0.81 ± 0.02
	SDCNet	0.93 ± 0.01	0.85 ± 0.01	0.92 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.82 ± 0.01
	DeepSynergy	0.87 ± 0.01	0.81 ± 0.01	0.91 ± 0.01	0.85 ± 0.01	0.83 ± 0.01	0.88 ± 0.01
	Matchmaker	0.88 ± 0.01	0.82 ± 0.01	0.91 ± 0.01	0.86 ± 0.01	0.83 ± 0.01	0.89 ± 0.01
Bliss	Jiang's method	0.92 ± 0.01	0.83 ± 0.01	0.94 ± 0.01	0.87 ± 0.01	0.81 ± 0.01	0.90 ± 0.01
	DeepDDS	0.92 ± 0.01	0.87 ± 0.01	0.93 ± 0.01	0.90 ± 0.01	0.88 ± 0.02	0.92 ± 0.02
	DTF	0.94 ± 0.01	0.89 ± 0.01	0.96 ± 0.01	0.91 ± 0.01	0.89 ± 0.01	0.94 ± 0.01
	SDCNet	0.96 ± 0.01	0.90 ± 0.01	0.97 ± 0.01	0.92 ± 0.01	0.91 ± 0.02	0.93 ± 0.01
	DeepSynergy	0.88 ± 0.01	0.83 ± 0.01	0.94 ± 0.01	0.89 ± 0.01	0.87 ± 0.01	0.91 ± 0.01
	Matchmaker	0.89 ± 0.01	0.83 ± 0.01	0.95 ± 0.01	0.89 ± 0.01	0.88 ± 0.01	0.89 ± 0.02
ZIP	Jiang's method	0.91 ± 0.01	0.84 ± 0.01	0.96 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.94 ± 0.02
	DeepDDS	0.92 ± 0.01	0.89 ± 0.01	0.95 ± 0.01	0.93 ± 0.02	0.91 ± 0.02	0.94 ± 0.02
	DTF	0.94 ± 0.01	0.88 ± 0.01	0.97 ± 0.01	0.92 ± 0.01	0.90 ± 0.02	0.94 ± 0.03
	SDCNet	0.95 ± 0.01	0.91 ± 0.01	0.98 ± 0.01	0.94 ± 0.01	0.92 ± 0.01	0.96 ± 0.01
	DeepSynergy	0.81 ± 0.02	0.86 ± 0.01	0.95 ± 0.01	0.90 ± 0.01	0.86 ± 0.01	0.94 ± 0.01
	Matchmaker	0.81 ± 0.02	0.86 ± 0.01	0.95 ± 0.01	0.92 ± 0.01	0.90 ± 0.01	0.94 ± 0.01
HSA	Jiang's method	0.90 ± 0.01	0.84 ± 0.01	0.96 ± 0.01	0.91 ± 0.01	0.84 ± 0.01	0.92 ± 0.01
	DeepDDS	0.89 ± 0.01	0.89 ± 0.01	0.96 ± 0.01	0.94 ± 0.01	0.92 ± 0.01	0.96 ± 0.02
	DTF	0.92 ± 0.01	0.89 ± 0.01	0.97 ± 0.01	0.93 ± 0.01	0.91 ± 0.01	0.96 ± 0.01
	SDCNet	0.94 ± 0.01	0.92 ± 0.01	0.98 ± 0.01	0.95 ± 0.01	0.93 ± 0.02	0.97 ± 0.01

Table 3. The performances of SDCNet and state-of-the-art methods on the ALMANAC dataset with various synergy types

Synergy type	Methods	AUC	ACC	AUPR	F1	Precision	Recall
	DeepSynergy	0.82 ± 0.06	0.57 ± 0.06	0.83 ± 0.08	0.23 ± 0.16	0.90 ± 0.17	0.14 ± 0.12
Loewe	Matchmaker	0.82 ± 0.06	0.54 ± 0.03	0.83 ± 0.08	0.13 ± 0.09	0.92 ± 0.14	0.07 ± 0.05
	Jiang's method	0.53 ± 0.06	0.50 ± 0.02	0.49 ± 0.05	0.36 ± 0.15	0.51 ± 0.04	0.31 ± 0.17
	DeepDDS	0.86 ± 0.03	0.68 ± 0.05	0.83 ± 0.03	0.51 ± 0.11	1.0 ± 0.0	0.35 ± 0.11
	DTF	0.87 ± 0.02	0.64 ± 0.06	0.88 ± 0.03	0.44 ± 0.09	0.99 ± 0.01	0.29 ± 0.07
	SDCNet	0.85 ± 0.04	0.75 ± 0.08	0.88 ± 0.03	0.67 ± 0.12	0.96 ± 0.03	0.52 ± 0.14
	DeepSynergy	0.70 ± 0.01	0.64 ± 0.01	0.70 ± 0.01	0.67 ± 0.01	0.64 ± 0.01	0.70 ± 0.02
	Matchmaker	0.70 ± 0.01	0.65 ± 0.01	0.71 ± 0.01	0.68 ± 0.01	0.64 ± 0.01	0.73 ± 0.02
Bliss	Jiang's method	0.80 ± 0.01	0.69 ± 0.01	0.75 ± 0.01	0.71 ± 0.02	0.72 ± 0.02	0.72 ± 0.02
	DeepDDS	0.82 ± 0.01	0.74 ± 0.01	0.81 ± 0.01	0.74 ± 0.02	0.75 ± 0.02	0.74 ± 0.03
	DTF	0.84 ± 0.01	0.75 ± 0.01	0.84 ± 0.01	0.76 ± 0.02	0.76 ± 0.02	0.77 ± 0.02
	SDCNet	0.86 ± 0.01	0.78 ± 0.01	0.86 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	0.78 ± 0.01
	DeepSynergy	0.74 ± 0.01	0.69 ± 0.01	0.67 ± 0.01	0.68 ± 0.01	0.62 ± 0.01	0.76 ± 0.01
ZIP	Matchmaker	0.79 ± 0.01	0.74 ± 0.01	0.71 ± 0.01	0.70 ± 0.02	0.68 ± 0.02	0.70 ± 0.01
	Jiang's method	0.86 ± 0.01	0.84 ± 0.01	0.77 ± 0.01	0.75 ± 0.02	0.77 ± 0.02	0.72 ± 0.02
	DeepDDS	0.90 ± 0.01	0.82 ± 0.01	0.85 ± 0.01	0.80 ± 0.02	0.82 ± 0.02	0.78 ± 0.02
	DTF	0.91 ± 0.01	0.83 ± 0.01	0.88 ± 0.01	0.80 ± 0.02	0.79 ± 0.02	0.82 ± 0.02
	SDCNet	0.93 ± 0.01	0.86 ± 0.01	0.92 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.84 ± 0.01
	DeepSynergy	0.77 ± 0.01	0.72 ± 0.01	0.63 ± 0.01	0.52 ± 0.01	0.66 ± 0.02	0.47 ± 0.02
	Matchmaker	0.78 ± 0.01	0.73 ± 0.01	0.64 ± 0.01	0.57 ± 0.01	0.61 ± 0.01	0.53 ± 0.02
HSA	Jiang's method	0.83 ± 0.01	0.80 ± 0.01	0.78 ± 0.01	0.71 ± 0.01	0.75 ± 0.01	0.67 ± 0.02
	DeepDDS	0.88 ± 0.01	0.84 ± 0.01	0.80 ± 0.01	0.75 ± 0.01	0.80 ± 0.01	0.70 ± 0.02
	DTF	0.89 ± 0.01	0.82 ± 0.01	0.82 ± 0.01	0.73 ± 0.02	0.81 ± 0.01	0.65 ± 0.01
	SDCNet	0.90 ± 0.01	0.85 ± 0.01	0.85 ± 0.01	0.76 ± 0.01	0.82 ± 0.02	0.71 ± 0.01

The performance of different methods across cell lines
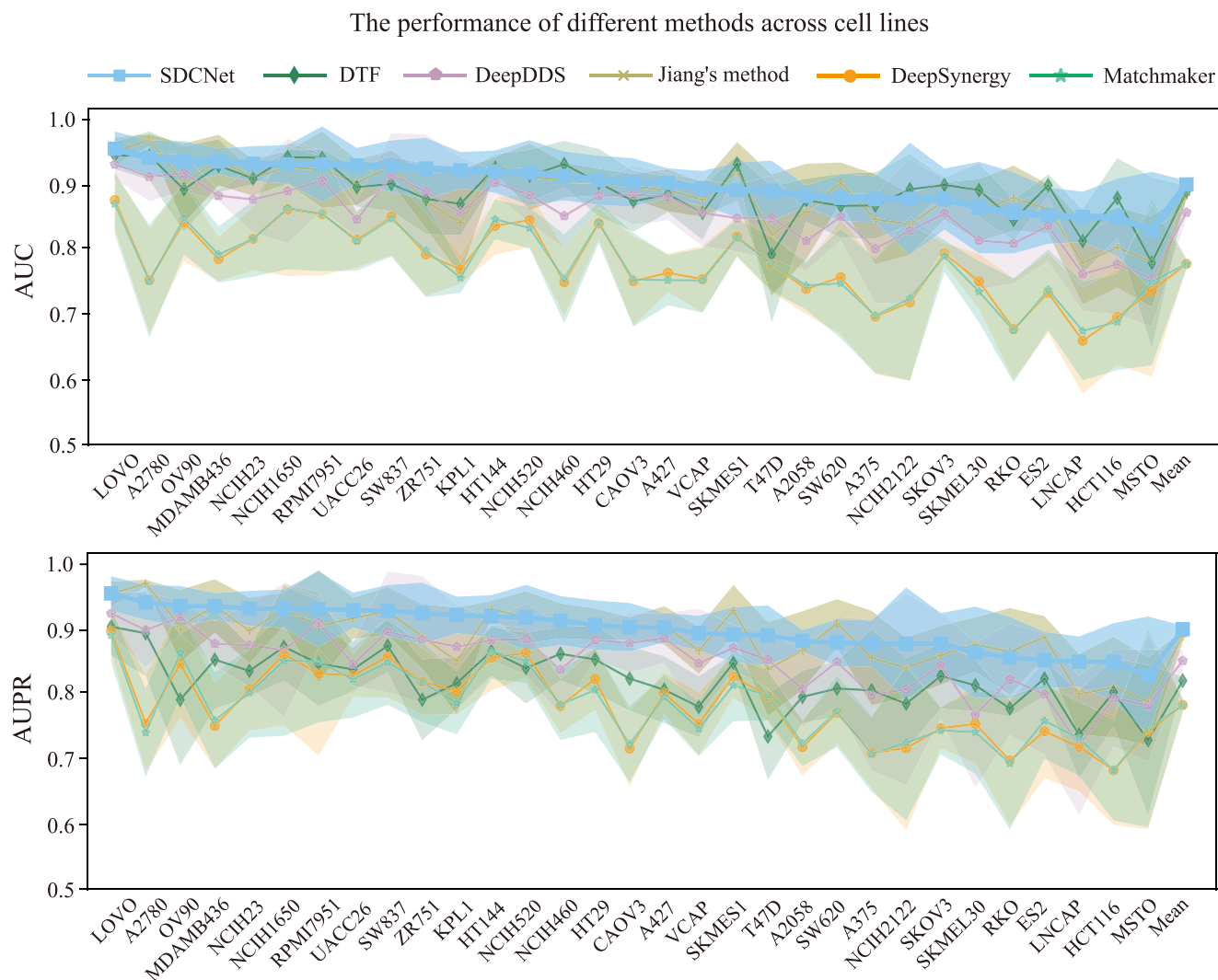
Figure 2. The AUC and AUPR values of different methods in each cell line on the O'Neil dataset (Loewe). The mean values of different folds are shown as solid lines. Error bars in terms of one standard deviation are shown as shaded areas.

Table 4. The performances of SDCNet based on different drug embeddings on the O'Neil dataset (Loewe)

| Methods	AUC	ACC	AUPR	F1	Precision	Recall
SDCNet-L1	0.90 ± 0.01	0.82 ± 0.01	0.90 ± 0.01	0.81 ± 0.01	0.83 ± 0.01	0.80 ± 0.02
SDCNet-L2	0.90 ± 0.01	0.83 ± 0.01	0.91 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.81 ± 0.02
SDCNet-L3	0.90 ± 0.01	0.83 ± 0.01	0.90 ± 0.01	0.83 ± 0.02	0.84 ± 0.01	0.81 ± 0.02
SDCNet-AVE	0.89 ± 0.01	0.82 ± 0.01	0.89 ± 0.01	0.81 ± 0.01	0.82 ± 0.01	0.80 ± 0.02
SDCNet-CON	0.91 ± 0.01	0.83 ± 0.01	0.91 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.82 ± 0.02
SDCNet	0.93 ± 0.01	0.85 ± 0.01	0.92 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.82 ± 0.01

that leverage the final drug embedding in different manners on the O'Neil dataset (Loewe), and the results are listed in Table 4. First, SDCNet-L1, SDCNet-L2 and SDCNet-L3 take the embeddings from different R-GCN layers as the final drug embeddings, respectively. Interestingly, similar performances in terms of various evaluation metrics are achieved using these models. Then, SDCNet-AVE averages the embeddings from different R-GCN layers and use them as the final drug embeddings, but the worst performance was obtained when using this model. In contrast, when SDCNet-CON concatenates these embeddings, the model's performance significantly increased compared with the previous models. Finally, SDCNet, which adopts attention mechanism to

manage these embeddings from different R-GCN layers, achieve the best performance among all attempts, with AUC, ACC, AUPR and F1 values of 0.93, 0.85, 0.92 and 0.83, respectively, indicating the effectiveness of the attention mechanism. Therefore, we apply attention mechanism to the embeddings from different R-GCN layers to obtain the final drug embeddings in the following study.

Influence of transfer learning

In this section, we aim to explore the influence of transfer learning on cell line-specific SDC prediction. Two types of SDCNet models are trained on the transfer dataset (Loewe). One model is the normal SDCNet model; the other model is the SDCNet model
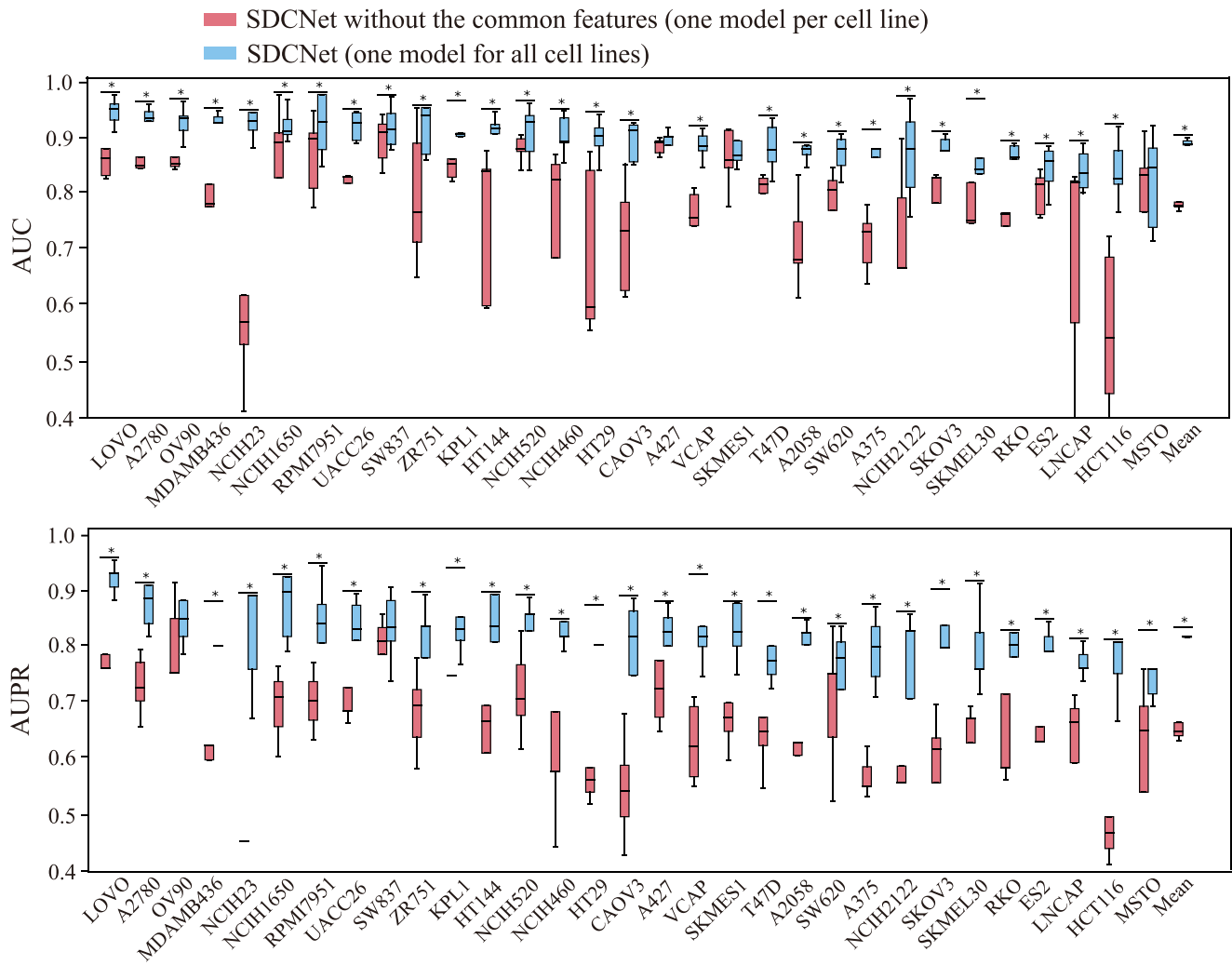
Figure 3. The AUC and AUPR values of SDCNet models with and without the common features in each cell line on the O'Neil dataset (Loewe). The P-values are calculated by Wilcoxon signed-rank test. * represents *P*-value <0.05.
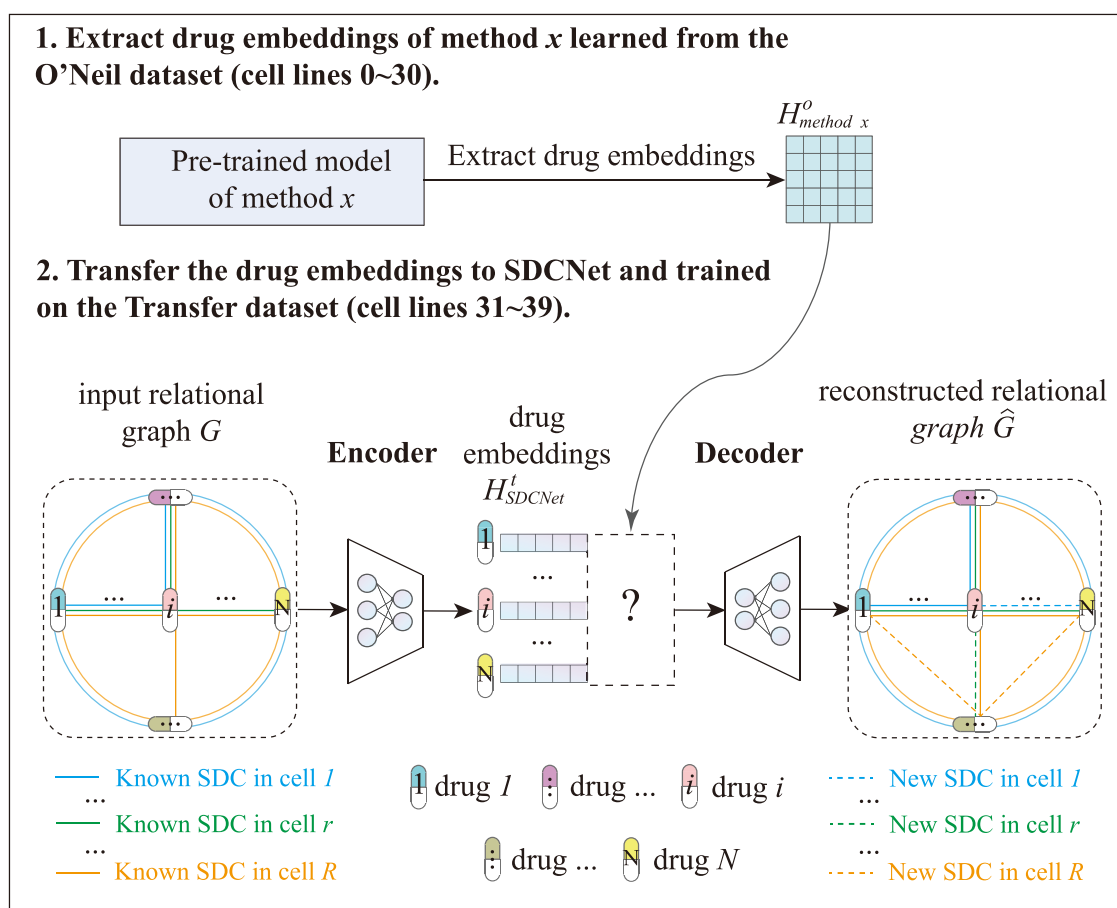
with the drug embedding, which learned from the O'Neil dataset (Loewe) (Figure 4). Then, the performances of these two models are compared to evaluate the effect of transfer learning.

To investigate the effectiveness of transfer learning for SDC-Net, we use the drug embedding extracted from the pre-trained SDCNet model. The SDCNet model with the drug embedding has significantly increased performance relative to the normal SDCNet model in terms of all evaluation metrics (Table 5). The above results demonstrate that SDCNet enables effective transfer of drug embedding to improve the prediction accuracy.
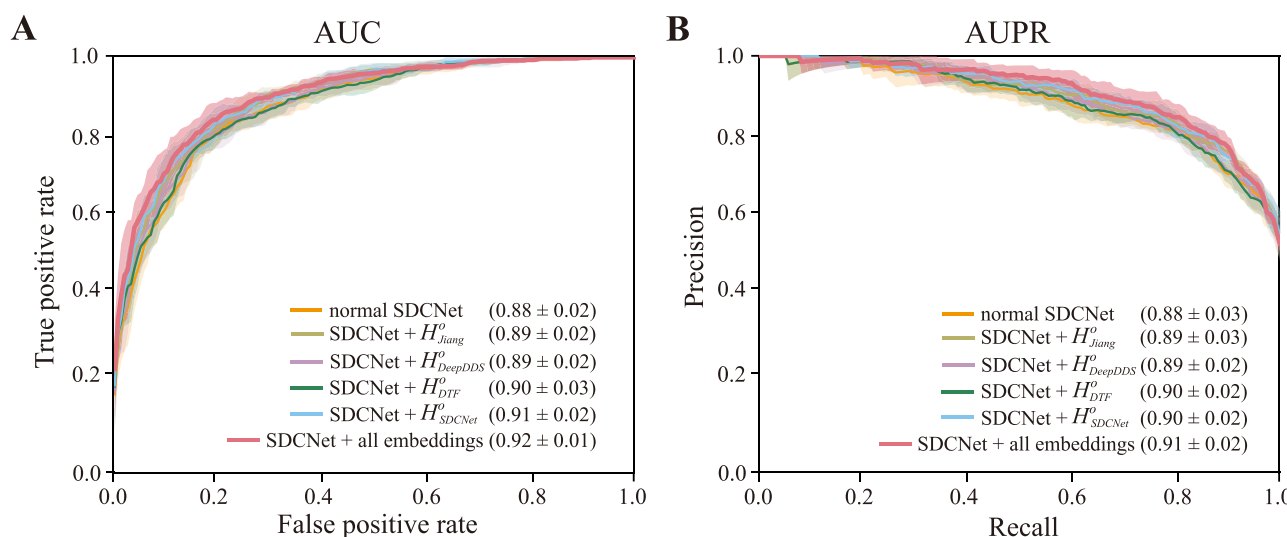
To further test whether other methods learn useful drug combination knowledge from the O'Neil dataset, we also use the drug embeddings extracted from the pre-trained models of Jiang's method, DTF and DeepDDS, respectively. The AUC and AUPR of the models show that the SDCNet models with the drug embeddings learned by these methods all performed better than the norm SDCNet model, indicating the effectiveness of transfer learning (Figure 5). Interestingly, we find that the SDCNet model that merges all of the drug embeddings learned by different methods (SDCNet, Jiang's method, DTF and DeepDDS) achieves the best performance among all the variants (Table 5), suggesting that different methods learn effective and complementary knowledge for drug combinations from the O'Neil dataset.

Case study

To test whether SDCNet can predict novel cell line-specific SDCs, we train the model based on the O'Neil dataset (Loewe) and predict the potential SDCs among previously untested drug combinations in each cell line. The drug combination with the highest probability score in each cell line indicated a high chance of showing synergistic effect in cancer treatment (Table 6). In total, 19 drug combinations achieve the highest probability score in different cell lines and most of their probability scores are higher than 0.95 (Supplementary Figure S3). We perform an in-depth literature survey and find evidence for many of these drug combinations. For example, etoposide and topotecan are both topoisomerase inhibitors, and their combination has been proven to show synergistic effect in ovarian cancer cell lines A2780 and SKOV3 [56–58]. We find that the probability score of etoposide and topotecan combination predicted by SDCNet achieves 0.99, ranking first place in the cell line SKOV3. We further check the prediction results of the combination in other ovarian cancer cell lines included in our dataset, namely, A2780, CAOV3, ES2 and OV90, which are 0.97, 0.97, 0.95 and 0.86, respectively. Although this combination does not obtain the highest probability score in these ovarian cancer cell lines, they are still in the top 10, indicating that the combination of etoposide and topotecan has

1. Extract drug embeddings of method *x* learned from the O'Neil dataset (cell lines 0~30).

2. Transfer the drug embeddings to SDCNet and trained on the Transfer dataset (cell lines 31~39).

Figure 4. The schematic of SDCNet model with the drug embeddings learned by specific method from the O'Neil dataset (Loewe). $H_{method\ x}^{dataset}$ means the drug embeddings learned by method x from the dataset. e.g. H_{SDCNet}^{t} and H_{SDCNet}^{o} represent the drug embeddings learned by method SDCNet from the Transfer dataset and the O'Neil dataset, respectively.

A AUC

normal SDCNet (0.88 ± 0.02)
SDCNet + H_{Jiang}^{o} (0.89 ± 0.02)
SDCNet + $H_{DeepDDS}^{o}$ (0.89 ± 0.02)
SDCNet + H_{DTF}^{o} (0.90 ± 0.03)
SDCNet + H_{SDCNet}^{o} (0.91 ± 0.02)
SDCNet + all embeddings (0.92 ± 0.01)

B AUPR

normal SDCNet (0.88 ± 0.03)
SDCNet + H_{Jiang}^{o} (0.89 ± 0.03)
SDCNet + $H_{DeepDDS}^{o}$ (0.89 ± 0.02)
SDCNet + H_{DTF}^{o} (0.90 ± 0.02)
SDCNet + H_{SDCNet}^{o} (0.90 ± 0.02)
SDCNet + all embeddings (0.91 ± 0.02)

Figure 5. The AUC and AUPR of the SDCNet models with different drug embeddings on the Transfer dataset (Loewe). H_{SDCNet}^{o}, H_{Jiang}^{o}, H_{DTF}^{o} and $H_{DeepDDS}^{o}$ represent the drug embeddings learned by method SDCNet, Jiang's method, DTF and DeepDDS from the O'Neil dataset (Loewe) separately. The averaged ROC curves are shown as solid lines. Error bars in terms of one standard deviation are shown as shaded areas.

highly synergistic potential for the treatment of ovarian cancer (Supplementary Table S11). Moreover, etoposide and 5-FU both have proven their values in separately treating melanoma, but their combination has not attracted much attention [59, 60].

Recently, their synergistic potential has been verified by the CCK8 test, an experiment for cellular proliferation, in the melanoma cell line A375 [55]. The predicted probability score for this combination by SDCNet is 0.99, the highest score in cell line A375.

Table 5. The performances of the SDCNet model with different drug embeddings on the Transfer dataset (Loewe)

Methods	AUC	ACC	AUPR	F1	Precision	Recall
normal SDCNet	0.88 ± 0.02	0.80 ± 0.01	0.88 ± 0.03	0.79 ± 0.02	0.79 ± 0.03	0.81 ± 0.02
SDCNet + H^o_{Jiang}	0.89 ± 0.02	0.81 ± 0.01	0.89 ± 0.03	0.80 ± 0.03	0.80 ± 0.01	0.82 ± 0.05
SDCNet + $H^o_{DeepDDS}$	0.89 ± 0.03	0.81 ± 0.02	0.89 ± 0.02	0.80 ± 0.02	0.80 ± 0.01	0.83 ± 0.03
SDCNet + H^o_{DTF}	0.90 ± 0.02	0.81 ± 0.02	0.90 ± 0.02	0.81 ± 0.02	0.81 ± 0.01	0.82 ± 0.03
SDCNet + H^o_{SDCNet}	0.91 ± 0.02	0.82 ± 0.02	0.90 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.83 ± 0.03
SDCNet + all embeddings	0.92 ± 0.01	0.83 ± 0.02	0.91 ± 0.02	0.83 ± 0.02	0.83 ± 0.03	0.83 ± 0.01

Table 6. The top SDC predicted by SDCNet in each cell line on the O'Neil dataset (Loewe)

Cell line ID	Cell line	Tissue	Drug 1	Drug 2	Probability	Publications (PMID)
0	A2058	Melanoma	SN-38	TOPOTECAN	0.97	NA
1	A2780	Ovarian	CARBOPLATIN	ETOPOSIDE	0.98	NA
2	A375	Melanoma	5-FU	ETOPOSIDE	0.99	35,062,018
3	A427	Lung	5-FU	TOPOTECAN	0.95	NA
4	CAOV3	Ovarian	CYCLOPHOSPHAMIDE	ETOPOSIDE	0.99	NA
5	ES2	Ovarian	PACLITAXEL	VINBLASTINE	0.99	NA
6	HCT116	Colon	SN-38	TOPOTECAN	0.96	NA
7	HT144	Melanoma	PACLITAXEL	SN-38	0.99	NA
8	HT29	Colon	CARBOPLATIN	ETOPOSIDE	0.99	NA
9	KPL1	Breast	5-FU	TOPOTECAN	0.99	14,583,785
10	LNCAP	Prostate	ETOPOSIDE	METFORMIN	0.98	NA
11	LOVO	Colon	DEXAMETHASONE	PACLITAXEL	0.89	NA
12	MDAMB436	Breast	ETOPOSIDE	SN-38	0.99	NA
13	MSTO	Lung	TOPOTECAN	VINORELBINE	0.99	18,096,059
14	NCIH1650	Lung	SN-38	TOPOTECAN	0.99	NA
15	NCIH2122	Lung	CYCLOPHOSPHAMIDE	ETOPOSIDE	0.98	NA
16	NCIH23	Lung	MITOMYCINE	SN-38	0.93	NA
17	NCIH460	Lung	ETOPOSIDE	PACLITAXEL	0.99	11,233,806
18	NCIH520	Lung	ETOPOSIDE	TOPOTECAN	0.99	15,893,010
19	OV90	Ovarian	5-FU	VINORELBINE	0.99	NA
20	RKO	Colon	SN-38	TOPOTECAN	0.99	NA
21	RPMI7951	Melanoma	SN-38	TOPOTECAN	0.99	NA
22	SKMEL30	Melanoma	ETOPOSIDE	PACLITAXEL	0.99	NA
23	SKMES1	Lung	SN-38	TOPOTECAN	0.99	NA
24	SKOV3	Ovarian	ETOPOSIDE	TOPOTECAN	0.99	16,412,499, 15,956,976
25	SW620	Colon	TOPOTECAN	VINBLASTINE	0.99	14,583,785
26	SW837	Colon	SN-38	TOPOTECAN	0.9	NA
27	T47D	Breast	5-FU	METFORMIN	0.98	NA
28	UACC62	Melanoma	SN-38	TOPOTECAN	0.99	NA
29	VCAP	Prostate	DOXORUBICIN	TOPOTECAN	0.98	14,583,785
30	ZR751	Breast	DEXAMETHASONE	ETOPOSIDE	0.98	NA

This combination is also included in the top 10 highest scores for other melanoma cell lines, including PRMI7951 (0.88), UACC62 (0.85) and SKMEL30 (0.92). Based on these results, it is believed that the prediction results of SDCNet are consistent with many previous studies, and SDCNet can accurately identify known cell line-specific SDCs and predict novel reliable SDCs.

Conclusions

In this paper, we propose a novel efficient method named SDCNet to predict potential SDCs for cancer treatment *in silico*. Compared with the existing computational methods, SDCNet learns and fuses the unique features of drug combinations in a specific cell line and their common patterns across different cell lines. Experiments on different datasets demonstrate that SDCNet is superior to state-of-the-art methods in predicting cell line-specific SDCs on

datasets from multi-cell lines. Additionally, SDCNet enables effective transfer of drug embedding to further improve the prediction accuracy.

However, our method still has some limitations and we will make more efforts to improve model performance. The greatest challenge for predicting the efficient SDCs through computational methods is the lack of sufficient data size, namely, the drug combinations that have been experimentally identified the synergy scores [61]. This also limited the performance of SDCNet, who focus on predicting the synergistic effects of drug combinations in specific cell lines. Since the common features of drug combination among cell lines can improve cell line-specific SDC prediction, we anticipate that SDCNet will obtain better performance if more data are available in the future. Another challenge in predicting SDCs is the biological interpretation of computational methods especially ML models [62]. Hidden prior knowledge is often vital to improve model performance and a better understanding of the

mechanism underlying SDCs [48]. For instance, Jiang et al. leveraged prior information from drug-target interaction network to improve prediction accuracy [10]. In the future, we are interested in incorporating more chemical/biological knowledge to develop more powerful prediction model and increase the interpretability of the model.

Key Points

- We present a novel encoder-decoder network named SDCNet to efficiently predict cell line-specific SDCs facilitating the discovery of rational combination therapies. It can learn and fuse the unique features of drug combinations in a specific cell line and their invariant patterns across different cell lines, and the common features can improve the prediction accuracy for each cell line.
- SDCNet enables effective transfer of deep drug embedding, which learned from other datasets to further improve the prediction accuracy.
- Experiments on different datasets demonstrate that SDCNet is superior to state-of-the-art methods in predicting cell line-specific SDCs.

Data availability

The drug combinations' synergy datasets are extracted from https://drugcomb.fimm.fi. The gene expression profiles of cell lines are derived https://sites.broadinstitute.org/ccle/, while the drugs SMILES are obtained from https://go.drugbank.com/. The implementation of SDCNet and the preprocessed data are available at https://github.com/yushenshashen/SDCNet.

Funding

References

1. Dupont CA, Riegel K, Pompaiah M, *et al*. Druggable genome and precision medicine in cancer: current challenges. *FEBS J* 2021;**288**:6142–58.
2. Jia J, Zhu F, Ma X, *et al*. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009;**8**: 111–28.
3. Rikkala PR, Jha SS, Pore D, *et al*. A review on drug combination strategy for pharma life cycle management. *J Biol Today's World* 2020;**9**:215.
4. Liu J, Gefen O, Ronin I. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science, 204* **202**(367):200.
5. Sicklick JK, Kato S, Okamura R, *et al*. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 2019;**25**:744–50.
6. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence* 2020;**2**:573–84.
7. Yin N, Ma W, Pei J, *et al*. Synergistic and antagonistic drug combinations depend on network topology. *PLoS One* 2014;**9**:93960.
8. Jin W, Stokes JM, Eastman RT, *et al*. Deep learning identifies synergistic drug combinations for treating COVID-19. *Proc Natl Acad Sci U S A* 2021;**118**:e2105070118.
9. Gerdes H, Casado P, Dokal A, *et al*. Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. *Nat Commun* 2021;**12**:1850.
10. Jiang P, Huang S, Fu Z, *et al*. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Comput Struct Biotechnol J* 2020;**18**:427–38.
11. Yang M, Jaaks P, Dry J, *et al*. Stratification and prediction of drug synergy based on target functional similarity. *NPJ Syst Biol Appl* 2020;**6**:16.
12. Pang K, Wan YW, Choi WT, *et al*. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics* 2014;**30**:1456–63.
13. Vakil V, Trappe W. Drug combinations: mathematical modeling and networking methods. *Pharmaceutics* 2019;**11**:208.
14. EILL F. Combination chemotherapy of acute leukemia and lymphoma. *JAMA* 1972;**7**:91–121.
15. Ter-Levonian AS, Koshechkin KA. Review of machine learning technologies and neural networks in drug synergy combination pharmacological research. *Research Results in Pharmacology* 2020;**6**:27–32.
16. Holbeck SL, Camalier R, Crowell JA, *et al*. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res* 2017;**77**:3564–76.
17. Sidorov P, Naulaerts S, Ariey-Bonnet J, *et al*. Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front Chem* 2019;**7**:509.
18. O'Neil J, Benita Y, Feldman I, *et al*. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016;**15**:1155–62.
19. Liu Y, Wei Q, Yu G, *et al*. DCDB 2.0: a major update of the drug combination database. *Database (Oxford)* 2014;**2014**:1–6.
20. Zagidullin B, Aldahdooh J, Zheng S, *et al*. DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019;**47**:W43–51.
21. Hu W, Liu B, Gomes J. Strategies for pre-training graph neural network. *ICLR* 2020;1–15.
22. Ianevski A, Giri AK, Gautam P, *et al*. Prediction of drug combination effects with a minimal set of experiments. *Nat Mach Intell* 2019;**1**:568–77.
23. Ianevski A, Giri AK, Aittokallio T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res* 2020;**48**:488–93.
24. Guvenc Paltun B, Kaski S, Mamitsuka H. Machine learning approaches for drug combination therapies. *Brief Bioinform* 2021;22:bbab293.
25. Chen H, Li J. DrugCom: synergistic discovery of drug combinations using tensor decomposition. *IEEE International Conference on Data Mining (ICDM)* 2018;**2018**: 899–904.
26. Preuer K, Lewis RPI, Hochreiter S, *et al*. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46.
27. Kuru HI, Tastan O, Cicek E. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**:2334–2344.
28. Liu Q, Xie L. TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol* 2021;**17**:1008653.

29. Sun Z, Huang S, Jiang P, *et al.* DTF: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics* 2020;**36**: 4483–9.

30. Yu Z, Huang F, Zhao X, *et al.* Predicting drug-disease associations through layer attention graph convolutional network. *Brief Bioinform* 2021;**22**:1–11.

31. Nguyen T, Le H, Quinn TP, *et al.* GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics* 2021;**37**:1140–7.

32. Wang J, Liu X, Shen S, *et al.* DeepDDS-deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2021;**23**:1–12.

33. Jaaks P, Coker EA, Vis DJ, *et al.* Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* 2022;**603**: 166–73.

34. Schlichtkrull M, Kipf TN, Bloem P, *et al.* Modeling relational data with graph convolutional networks . *European semantic web conference.* Springer, Cham, 2018: 593–607.

35. Wang J, Li J, Yue K, *et al.* NMCMDA: neural multicategory MiRNA-disease association prediction. *Brief Bioinform* 2021;**22**:1–11.

36. Peng J, Wang Y, Guan J, *et al.* An end-to-end heterogeneous graph representation learning-based framework for drug-target interaction prediction. *Brief Bioinform* 2021;**22**:1–9.

37. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**:457–66.

38. Zheng S, Aldahdooh J, Shadbahr T, *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;**49**:174–84.

39. Greco WR, Bravo G, Parsons JC. The search for synergy-a critical review from a response surface perspective. *Pharmacol Rev* 1995;**47**:331–85.

40. Bliss CI. The toxicity of poisons applied jointly. *Ann Appl Biol* 1939;**26**:585–615.

41. Yadav B, Wennerberg K, Aittokallio T, *et al.* Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Comput Struct Biotechnol J* 2015;**13**:504–13.

42. Berenbaum MC. What is synergy? *Pharmacol Rev* 1989;**41**:93–141.

43. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:1074–82.

44. Capecchi A, Probst D, Reymond JL. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *J Chem* 2020;**12**:43.

45. Zagidullin B, Wang Z, Guan Y, *et al.* Comparative analysis of molecular fingerprints in prediction of drug combination effects. *Brief Bioinform* 2021;**22**:1–15.

46. Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. *J Chem Inf Model* 2015;**55**:2324–37.

47. Vashishth S, Sanyal S, Nitin V. Composition-based multi-relational graph convolutional networks. *ICLR* 2020;1–14.

48. Wang Z, Li H, Guan Y. Machine learning for Cancer drug combination. *Clin Pharmacol Ther* 2020;**107**:749–52.

49. ZHou K, Huang X, Li Y. Towards deeper graph neural networks with differentiable group normalization, *Advances in neural information processing systems*, 2020;**33**:4917–4928.

50. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. *IN NIPs* 2016;**30**:3844–52.

51. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010*, 2010, 249–56.

52. Kingma DP, Ba JL. Adam: a method for stochastic optimization. *ICLR* 2015;1–9.

53. Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout-a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;**15**:1929–58.

54. Smith LN. Cyclical learning rates for training neural networks. *WACV* 2017;464–472.

55. Lin W, Wu L, Zhang Y, *et al.* An enhanced cascade-based deep forest model for drug combination prediction. *Brief Bioinform* 2022;**23**:1–18.

56. Lee YS, Lee TH. Effect of Topotecan in Combication with other antitumor drugs in vitro. *Korean J Gynecol Oncol Colposc* 2019;**11**(1): 83–90.

57. Penson RT, Seiden MV, Matulonis UA, *et al.* A phase I clinical trial of continual alternating etoposide and topotecan in refractory solid tumours. *Br J Cancer* 2005;**93**:54–9.

58. Reck M, Groth G, Buchholz E, *et al.* Topotecan and etoposide as first-line therapy for extensive disease small cell lung cancer: a phase II trial of a platinum-free regimen. *Lung Cancer* 2005;**48**: 409–13.

59. Ryan RF, Krementz ET, Litwin MS. A role for topical 5-fluorouracil therapy in melanoma. *J Surg Oncol* 1988;**38**(4): 250–6.

60. Rudolf K, Cervinka M, Rudolf E. Cytotoxicity and mitochondrial apoptosis induced by etoposide in melanoma cells. *Cancer Invest* 2009;**27**:704–17.

61. Wang Z, Li H, Carpenter C, *et al.* Challenge-enabled machine learning to drug-response prediction. *AAPS J* 2020;**22**:106.

62. Menden MP, Wang D, Mason MJ, *et al.* Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun* 2019;**10**:2674.