DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations

Jinxian Wang[a], Wenhao Zhang[b], Siyuan Shen[c], Lei Deng[a,c,*], Hui Liu[b,*]

[a]*School of Computer Science and Engineering, Central South University, Changsha 410075, China.*
[b]*School of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China.*
[c]*School of Software, Xinjiang University, Urumqi 830008, China.*

Abstract

Drug combination therapy becomes promising method in the treatment of cancer. However, the number of possible drug combinations toward cancer cell lines is too large, and it is challenging to screen synergistic drug combinations through wet-lab experiments. Therefore, the computational screening has become an important way to prioritize drug combinations. Graph attention network has recently shown strong performance in screening of compound-protein interactions, but it has not been applied to the screening of drug combinations. In this paper, we proposed a deep learning model (DeepDDS) based on graph neural networks and attention mechanism to identify drug combinations that can effectively inhibit the viability of specific cancer cell line. The graph representation of drug molecule structure and gene expression profiles is taken as input to predict the synergistic effects of drug combinations. We compare DeepDDS with traditional machine learning methods (random forest, support vector machine) and other deep learning methods (DeepSynergy, DTF) on the same data set. Our experimental results show that DeepDDS achieved best performance by the AUC value 0.93. Also, on an independent test set released by AstraZeneca, DeepDDS is superior to other comparative methods by 12.2% higher than the suboptimal method. We believe that DeepDDS is a effective tool that can prioritize synergistic drug combinations.

Keywords: Drug combination, Graph attention network, compound-protein interactions, deep learning

Introduction

Traditional and modern medicine has always taken advantage of the combined use of several active agents to treat different diseases. Compared with single-drug therapy, the combinatorial drugs instead of monotherapy can improve efficacy[1], reduce side effects[2] and overcome drug resistance[3, 4].Drug combinations are increasingly used to treat a variety of complex diseases, such as hypertension[5], infectious diseases[6], and cancer [7, 8].

*Corresponding author.
Email address: hliu@cczu.edu.cn (Hui Liu)

For example, triple-negative breast cancer is a malignant tumor with strong invasiveness, high metastasis rate, and poor prognosis. Studies have shown that lapatinib or rapamycin alone has no significant effect, but their combined treatment can significantly increase the apoptosis rate of triple-negative breast cancer cells[9]. However, drug combination may also cause antagonistic effect and aggravate the disease[10]. Therefore, it is crucial to accurately discover synergistic drug combinations for specific disease types to improve the treatment of patients.

The traditional screening based on clinical trials are limited to a tiny number of drugs[11], far from meeting the urgent need of anticancer drugs. Due to the huge number of possible drug combinations, traditional screening is cost-consuming and impractical. With the development and application of high-throughput drug screening technology, people can simultaneously carry out large-scale combinatorial drug screening on hundreds of cancer cell lines[12–14]. Torres et al. utilized yeast to screen large drug combinations and provided a method to identify preferential drug combinations for further testing in human cells[15]. In despite of high degree of genomic correlation between the original tumor and the derived cancer cell line, *in vitro* experiments of high-throughput drug screening still cannot accurately capture the mode of action of drug molecules *in vivo*[16]. Microcalorimetry screening[17] and genetically encoded fluorescent sensors[18] have also been used to screen effective antimicrobial combinations in *in vivo* treatment. However, these technique all require skilled operations and complicated experimental procedures.

In recent years, a few datasets of drug sensitivities to cancer cell lines greatly increase, such as Cancer Cell Line Encyclopedia (CCLE)[19] and Genomics of Drug Sensitivity in Cancer (GDSC), which contains a compilation of gene expression, chromosomal copy number and massively parallel sequencing data, as well as drug sensitivity to hundreds of human cancer cell lines. Meanwhile, several large-scale data resource of drug combinations have been proposed. For example, DrugCombDB[20] has more than 6,000,000 quantitative drug dose responses, from which they calculated multiple synergy scores to determine the overall synergy or antagonism of the drug combination. Recently, O'Neil et al. released a large-scale high-throughput drug pair synergy study, which included more than 20,000 drug pair synergy scores[21]. The famous pharmaceutical company AstraZeneca[22] also released the latest drug pair collaboration experiments, which includes 11,576 experiments of 910 drug combinations to 85 cancer cell lines with genomic characteristics. The above-mentioned database and dataset provides reliable data support for the development of computational study. As a result, many computational methods have been proposed to explore the vast space of combinatorial drug for predicting synergistic combinations. For example, traditional machine learning methods, such as support vector machine (SVM) and random forest, predict the maximal antiallodynic effect of a new derivative of dihydrofuran-2-one (LPP1) used in combination with pregabalin (PGB) in the streptozocin-induced neuropathic pain model in mice[23, 24]. Recently, the deep learning model DeepSynergy[25] uses the chemical information of drugs and the genomic features of disease to predict drug pairs with synergistic effects. We can study the synergy of anticancer drugs from multiple aspects. From the perspective of the gene regulation level[26], a set of optimal control nodes can be identified for perturbation on the disease gene regulatory network. Also, system pharmacology[27]

characterize the disease and drug action mechanisms at the system level, via connectivity mapping and network concentration analysis[28, 29]. Such multiscale understanding can enable precision medicine by promoting the rational development of combination therapy at individual patients for specific cancer.

Many researchers have successfully applied SMILES to characterize chemical properties of drugs. For example, Gao et al. used the drug descriptors based on the SMILES to predict drug synergy. Liu et al. regarded the SMILES code as a string and directly input into a convolutional neural network[30]. In this paper, we propose a deep learning model, DeepDDS (Deep Learning for Predict Drug-Drug Synergy), to predict the synergistic effect of drug combinations. We employ graph embedding of drugs and gene expression profiles as input to deep network. For the graph representation of drug chemical structure, the vertices are atoms, and the edges are chemical bonds. Next, a graph convolutional network and attention mechanism is used to compute the embedding vector of drug. By integration of genomic features, DeepDDS can capture the important feature of the data to distinguish drug synergy to specific cancer cell lines. We compare the prediction results of DeepDDS to other latest deep learning and machine learning methods on two public data set[21]. The comparative methods include DeepSynergy[25], DTF[31], support vector machine[23], and random forest[24]. Overall, we found that DeepDDS can accurately predict drug synergy to specific cancer cell line, and outperform other methods.

Materials and methods

Data source

The drug combinations were obtained from a recently released large-scale oncology screen dataset by Merck&Co[21], where the viability of cells from tens of cancer cell lines treated with over ten thousands of drug combinations was screened. The Loewe Additivity score[32], a synergy score that can define the attribute (synergistic or antagonistic) of drug combination based on the full dose-response screening matrix, was calculated for each drug pair using the batch processing mode of Combenefit[33]. Specifically, a combination with a score above 0 is classified as synergistic, a combination below 0 is antagonistic. Of note, replicates for one combination were tested in the original dataset. Therefore, the loewe scores of the replicates for the same drug combination versus cell line were averaged, yielding a more efficient benchmark training set that contains 22,737 unique drug-pair-cell-line combinations, covering 38 anticancer drugs (14 experimental drugs and 24 approved) and 39 human cancer cell lines from 7 different tissue types. According to the final Loewe scores, we manually classified the drug combinations as synergistic or antagonistic.

The gene expression data of cancer cell lines were derived from Cancer Cell Line Encyclopedia[19], an independent large-scale platform that makes the effort to characterize genomes, mRNA expression, and anti-cancer drug dose responses across cell lines. By retrieving Drugbank[34], we can obtain the SMILES (Simplified Molecular Input Line Entry System)[35] of the drugs, which are afterwards converted to graphs.

Pipeline of DeepDDS

In this paper, based on the combination of GNN and conventional CNN, we proposed a novel method, referred to as DeepDDS, to predict the synergistic effect of drug combinations. Figure 1 illustrates the schematic diagram of our proposed model. The chemical molecular structure map is non-Euclidean data and does not have translation invariance, from which the convolution kernel cannot extract the structural information, thus the CNN is inadequate for such type of data. Contrarily, DeepDDS can form a receiving field in a graph where the data points are not arranged in a Euclidean grid. We also applied the attention mechanism in the chemical molecular structure diagram. To verify the performance of attention mechanism, we designed two models, refer to DeepDDS (GAT), which uses an attention-based graph convolutional neural network, and DeepDDS (GCN), which uses a graph convolutional neural network, respectively.

Drug representation based on GCN and GAT

In this paper, by using the open-source chemical informatics software RDKit[36], we converted the SMILES into a molecular graph, where the vertices are atoms and the edges are chemical bonds. To describe a node in the graph, we used a set of atomic features adapted from DeepChem[37]. Specifically, each node is a multi-dimensional binary feature vector expressing five pieces of information: the atom symbol, the number of adjacent atoms, the number of adjacent hydrogens, the implicit value of the atom, and whether the atom is in an aromatic structure. The application of graph convolution to the molecular structure of compounds to obtain the characteristic information of the compound is the content of many kinds of research and has been achieved extensive research results[38].

A graph for a given drug is represented as $G = (V, E)$, where V is the set of N nodes that represented by a C-dimensional vector, and E is the set of edges represented as an adjacency matrix A. In a molecule, $v_i \in V$ is the ith atom and $e_{ij} \in E$ is the chemical bond between the ith and jth atoms. Message Passing that describes the learning process, is the essence of GNN, as shown in 1.

$$\mathbf{X}_i^{(k)} = \gamma^{(k)}(\mathbf{X}_i^{(k-1)}, \square_{j \in N(i)} \phi^{(k)}(\mathbf{X}_i^{(k-1)}, \mathbf{X}_j^{(k-1)}, \mathbf{e}_{i,j})) \tag{1}$$

where $\mathbf{x}$ is node embeddings, ϕ is message function, $\square$ is aggregation function, γ is update function, if the edges in the graph has no feature than connectivity , $\mathbf{e}$ is essentially the edge index of the gragh. The vector representation is formulated as follows:

$$\mathbf{X}_i^{(k)} = \sum_{j \in N(i) \cup \{i\}} \frac{1}{\sqrt{deg(i)} \cdot \sqrt{deg(j)}} \cdot (\Theta \cdot x_j^{(k-1)}) \tag{2}$$

where Θ is the weight matrix (the parameter to be updated in the neural network), $\mathbf{X}_i^{(k)}$ is the eigenvector of the kth iteration of node i, $deg(i)$ is the degree of node i, and $N(i)$ is the set of all neighbor nodes of node i.

graph convolutional network (GCN)

The input of multilayer graph convolutional network (GCN) is a node feature matrix $X \in \mathbb{R}^{N \times C}$ ($N = |V|$, C: the number of features per node) and an adjacency matrix $A \in \mathbb{R}^{N \times N}$. A layer-wise convolution operation in matrix can be approximated as follows:

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta \tag{3}$$

where the node-level output $Z \in \mathbb{R}^{N \times F}$ (F is the number of output features per node), $\Theta \in \mathbb{R}^{C \times F}$ (F is the number of filters or feature maps) is the matrix of filter parameters.

The normalized convolution propagation rules is formulated as follows:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \tag{4}$$

where $\tilde{A} = \tilde{A} + I_N$($I_N$ is the identity matrix) is the adjacency matrix of the undirected graph with added self-connections, $\tilde{D}_{ii} = \sum_i \tilde{A}_{ii}$; $H^{(l+1)} \in \mathbb{R}^{N \times C}$ is the matrix of activation in the lth layer, $H^{(0)} = X$, σ is an activation function, and W is learnable parameters. Finally, the max-pooling layer is used to obtain the representations generated by the graph neural networks[39].

Graph attention network (GAT)

Unlike GCN, the graph attention network (GAT) proposes an attention-based architecture to learn hidden features of nodes in a graph by applying a self-attention mechanism. The GAT architecture is built from the graphics attention layer. We use the node set of the graph as the input and linearly transform each node through the weight matrix W. The attention coefficient between each input node i and its first-order j neighbor in the graph is calculated as follows:

$$X_i^{'} = \alpha_{i,i} \Theta X_i + \sum_{j \in N(i)} \alpha_{i,j} \Theta X_j \tag{5}$$

The attention coefficients $\alpha_{i,j}$ can be computed as follows:

$$\alpha_{i,j} = \frac{exp(ReLU(a^T[\Theta X_i || \Theta X_j]))}{\sum_{k \in N(i) \cup \{i\}} exp(ReLU(a^T[\Theta X_i || \Theta X_k]))} \tag{6}$$

where Θ is weight matrix, a is attention coefficient vector, T is the corresponding transpose, and $ReLU$ is a Non-linear activation function, when x is negative, y is equal to 0. The 'softmax' function is introduced to normalize all neighbor nodes j of i for easy calculation and comparison.

Cell line representation based on DNN

To alleviate the imbalance between the features of drugs and cell lines, we further selected the significant genes according to Genomics of Drug Sensitivity in Cancer (GDSC)[40]. The LINCS project provides a set of about 1000 carefully chosen genes, referred to as 'Landmark gene set', which can capture 80% of the information[41] in the Connectivity Map (CMap)

data ([1]). The intersection between the original gene expression and the Landmark set was chosen. We subsequently used the gene annotation information in the Cancer Cell Line Encyclopedia (CCLE)[19] and the GENCODE annotation database[42] to remove the redundant data, as well as the transcripts of non-coding RNA. Finally, we selected **954** genes from raw expression as the dimension-reduced input of a deep neural network. In particular, our gene expression profile data comes from cell lines that have not been treated with the compound.

Predicting the synergistic effect of drug combinations versus cell lines

We proposed DeepDDS as an end-to-end binary classification model. Upon obtaining the embedding for drugs through GAT or GCN, and the embedding for cell lines through DNN, the separate embeddings are concatenated as the input of a series of fully-connected layers. The probability of the synergistic effect (label) for the combination can be achieved by activating the output of the last hidden layer, as follows:

$$p_t = softmax\left(W_{out} \cdot a^l + b_{out}\right) \tag{7}$$

where p_t is the probability of t, W_{out} and b_{out} are the weight matrix and bias vector, a^l are the embedding features learned by previous layers. as follows:

$$a^l = \sigma(W^l a^{l-1} + b^l) \tag{8}$$

Where l is the number of hidden layers, W and b are the matrices corresponding to all hidden layers and output layers, bias vector, $a^1 = concat(R_{drug1}, R_{drug2}, R_{cellline})$ is the input value vector.

Given a set of combinations with labels, we adopted the cross-entropy as the loss function to train the model, companying the aim to minimize the loss during the training process, which is formulated as follows:

$$F = min\left(-\sum_{i=1}^{N} logP_{t_i} + \frac{2}{\lambda}\|\Theta\|\right) \tag{9}$$

where Θ is the set of all weight matrices and bias vectors involved in the model, N is the total number of samples in the training dataset, t_i is the ith label, and λ is an L2 regularization hyper-parameter. Then, we use backpropagation to train Θ .

Result

Hyperparameter setting

The real architecture of DeepDDS was determined by the hyperparameter selection. DeepDDS-GAT has a spindle-shaped structure with three hidden layers. The first layer

[1]http://support.lincscloud.org/hc/en-us/articles/202092616-The-Landmark-Genes

has 780 units, the second layer has 512 units and the third layer has 128 units. Based on multi-head attention (head = 10), multiple independent attentions is calculated to prevent overfitting. The learning rate (5×10^{-5}) and the dropout rate of the input layer and hidden layer are set to 0.2. Use ELU and ReLU activation functions separately after the two GAT layers at DeepDDS-GAT. For DeepDDS-GCN, it also has similar layer structure, but only ReLU is used as activation function. We use a deep neural network (DNN) to obtain the hidden features of the cell line, and the hidden layers in the deep neural network structure has 2048, 1024 and 512 neural units, respectively.

Performance comparison

We compared DeepDDS with several state-of-the-art methods, includes:

- **DeepSynergy.** DeepSynergy[25] uses molecular chemistry and cell line genome information as input, and uses three different types of input normalization. A cone layer is used in a neural network (DNN) to simulate drug synergy and finally predict the synergy score. We use the same input data as DeepSynergy.

- **Random Forests(RF).** Random forest contains many classification trees. Put the input features on each tree in the forest to classify new objects. Each decision tree is a classifier, so for an input sample, N trees will have N classification results, we choose a moderate value[43]. Finally, the random forest integrates all the classification voting results, and uses the category with the most votes as the final output category.

- **Gradient Boosting Machines(GBM).** Train the Gradient Boosting Machines by continuously adjusting the decision tree, Boosting and other model parameters. Construct the new base-learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble[44].

- **Support Vector Machines(SVM).** We developed an SVM-based machine learning model[45] using the same input features as DeepDDS. Continuously optimize the SVM model's prediction performance through grid search so that C (penalty coefficient) and g (gamma: determine the distribution of the data mapped to the new feature space) take values within a specific range. Use the K-CV method to obtain the validation set's classification accuracy under the combination of c and g, and finally get the group of c and g with the highest validation classification accuracy of the training set as the best parameters.

To further benchmark the predictive power of DeepDDS, we provide typical performance measures for the classification tasks, including area under the receiver operator characteristics curve (ROC AUC), area under the precision recall curve (PR AUC), accuracy (ACC), balanced accuracy (BACC), precision (PREC), sensitivity (TPR) and Cohen's Kappa. Table 1 shows these performance measures of DeepDDS and counterpart methods. According to the experimental definition, all synergy scores higher than zero indicate synergy. However, drug combinations that exhibit a high degree of synergy are attractive candidates

for clinical research. Since there are many additive combinations in the training data (ie synergy score around 0), we choose an appropriate threshold to binarize the drug synergy score. Synergistic combinations with a measurement score higher than 10 are classified as positive, and synergistic combinations with a measurement score less than 0 are classified as negative, making the ratio of positive and negative samples close to 1:1. DeepDDS-GAT, tends to make more conservative predictions, its performance measures of ROC AUC, PR AUC, ACC, BACC, PREC, TPR, TNR and Kappa are 0.93, 0.93, 0.85, 0.85, 0.85, 0.85, 0.85 and 0.71, respectively. The performance of DeepDDS was better than other model in terms of all these performance measures. Although DeepSynergy and DTF exhibited obtain relatively high ACC metrics, our model achieve outperforms other methods in other performance metrics.

Prediction analysis

We analyzed the correlation between the predictive score and the true synergy score of the drug pairs with the top 100 synergy scores (Supplementary Table S1) in DeepDDS-GAT, and the prediction accuracy reached 0.98. We analyzed the correlation between the overall predictive score and the synergy score, and found that the higher the synergy score, the higher the predictive score. Similarly, we also analyzed the correlation between the binary prediction score and the synergy score of the 100 drug pairs with the lowest synergy score, and the correlation coefficient also reach 0.83 (Supplementary Table S2).

Further, we verified the prediction upon different scheme of input features. In fact, drug A-drug B and drug B-drug A are regarded as two different input schemes. Figure 2 shows prediction results of different scheme of input features by DeepDDS (GAT) and DeepDDS (GCN), respectively. It can be found that most predicted values clustered at the top or bottom. The Pearson correlation coefficients predicted on the GAT and GCN models can reach 0.833 and 0.829, respectively. It can prove that our model is insensitive to the input scheme of input features of drug combinations. Besides, the five-fold cross-validated ROC AUC and PR AUC obtained by drug A-drug B and drug B-drug A can both reach or be close to 0.93. For example, for the cell line T37D, the drugs BEZ-235 and MK-8669, DASATINIB, LAPATINIB, GELDANAMYCIN, PD325901, ERLOTINIB, MK-4541, TEMOZOLOMIDE, VINORELBINE, ABT-888, etc. all have a high experimental synergy scores (Loewe>100). Expectedly, the prediction scores of these drug pairs also have prior ranks among all candidate drug pairs.

Evaluation on independent test set

To further demonstrate the performance of our method, we use Merck&Co drug combination dataset [21] as the training set to learn the DeepDDS model, and an independent test set released by AstraZeneca[22] is used to evaluate the performance of DeepDDS and other comparative methods. The independent test set contains 668 unique drug-pair-cell-line combinations, covering 58 drugs (Supplementary Table S3) and 24 cell lines (Supplementary Table S4).

Our prediction performance with the competitive model on the independent test set is shown in Table 2. It can be seen that our model performance is better than the competitive

model in all performance indicators. DeepDDS (GAT), DeepDDS (GCN), and DeepSynergy respectively predicted 431, 402, and 317 drug pairs in independent test sets. We found that the three models correctly predicted 162 groups ((99 drug pairs are synergistic, 63 drugs are antagonistic)) of the same drug pairs (Supplementary Table S5), accounting for DeepDDS (GAT), DeepDDS (GCN), and DeepSynergy predict 38% (162/421), 40% (162/402), and 51% (162/317). The confusion matrices of the three models are shown in Figure 3.

Predicting novel synergistic combinations

The above analysis results show that the DeepDDS model achieve high performance, which means that we can reliably use the model to predict novel synergistic combination. We use O'Neil drug combination dataset to train our model and selected 668 completely novel drug pairs, from AstraZeneca[22], for discovery exploration.

We found that a number of predicted synergistic combinations are consistent with the observations in previous studies. For example, crafter et al. found that the combination of signal inhibitors **AZD8931** and **AZD5363** can limits AKT inhibitor induced feedback and enhances antitumour efficacy in HER2-amplified breast cancer models[46]. The combination of **GDC-0941** and **AZD6244** delayed tumor formation in the setting of prevention and extended survival when used to treat advanced tumors[47], although no durable responses were observed by Anchez et al. Also, the experimental results of other studies confirmed the effectiveness of our predicted synergistic drug pairs, such as **AZD2014** and **AZD6244** in Bladder[48], **GDC-0941** and **MK2206** in breast[49] and **Afatinib** and **MK2206** in breast[50] etc.

Attention mechanism reveals important chemical substructure

We further explore the implications of attention mechanism in revealing the important chemical substructure. In the GAT module, DeepDDS convolves drug molecules based on the attention mechanism. Unlike GCN, GAT absorb different weights from each neighbor node so that each node capture the information of neighboring nodes. Similarly, each neuron is connected to the neighborhood upper layer through a set of learnable weights in the GAT network. As a result, the atom feature representation actually include the functional information that exists in the substructure of the compound centered on the atom. From the structural point of view of the graph, these information includes substructure formal charge, the number of connected hydrogens, and undiscovered functional structure information. For example, a dose-response experiments showed that MEK inhibitor AZD6244 (Selumetinib) and mTORC1/2 inhibitor AZD2014 (Vistusertib) can effectively inhibit mTORC1 and S6 ribosomal protein[48] in 96 hours. As shown in the figure 4, AZD2014 after GAT and max-pooling obtains the C atom characteristics in the methyl group in the substructure A region. The same is true for AZD6244. We speculate that these two substructures may positively affect the synergy of the pair of drugs.

Discussion and Conclusion

For both GAT and GCN models, the effect of GAT in the five-fold cross-validation experiment is slightly better than that of GCN, and GAT is marginally higher in PR AUC

and KAPPA indicators. In the independent test set experiment, the overall performance of GCN will be slightly better than GAT, but the performance of both is unstable. In general, the predictive values of all method are low, so new drugs and promotion cannot be reliably predicted. We believe that the lower prediction performance is due to the smaller number of training samples. Specifically, all models can only train 38 drugs and 39 cell lines, and the space for possible drugs is much larger. And we have observed that drug pairs with high prediction scores often have higher synergy scores. Therefore, it is reasonable to predict a small number of drug combinations as highly synergistic pairs. It is worth noting that, in the GAT and GCN modules of the DeepDDS model that we do not yet fully understand the specific chemical physical meaning of the weights between the atoms in the molecular graph and their adjacent atoms. In the future, we are interested in studying the connections between atoms in the model to incorporate more information resources into the DeepDDS model to improve the model interpretability and predictability.

We have proposed a novel method DeepDDS to predict the synergy of drug combinations for cancer cell lines with high accuracy. We have demonstrate that DeepDDS achieve state-of-the-art performance in a cross-validation setting with an independent test set. Our performance comparison experiments showed that DeepDDS performs better than other comparative methods. Thanks to the increase in the size of the data set, DeepDDS can be further improved. We believe that our method can be applied to other fields where drug combinations play an essential role, such as antiviral[51], antifungal[52] and multi-drug synergy prediction[53]. Overall, our findings indicate that DeepDDS is a effective tool to predict synergistic drug combinations.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found at Supplementary file 1.

CONFLICTS OF INTEREST

The authors declare no competing interests.

References

[1] P. Csermely, T. Korcsmáros, H. J. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review, Pharmacology & therapeutics 138 (3) (2013) 333–408.

[2] S. Zhao, T. Nishimura, Y. Chen, E. U. Azeloglu, O. Gottesman, C. Giannarelli, M. U. Zafar, L. Benard, J. J. Badimon, R. J. Hajjar, et al., Systems pharmacology of adverse event mitigation by drug combinations, Science translational medicine 5 (206) (2013) 206ra140–206ra140.

[3] J. A. Hill, R. Ammar, D. Torti, C. Nislow, L. E. Cowen, Genetic and genomic architecture of the evolution of resistance to antifungal drug combinations, PLoS Genet 9 (4) (2013) e1003390.

[4] A. D. Verderosa, R. Dhouib, Y. Hong, T. K. Anderson, B. Heras, M. Totsika, A high-throughput cell-based assay pipeline for the preclinical development of bacterial dsba inhibitors as antivirulence therapeutics, Scientific Reports 11 (1) 1–13.

[5] T. D. Giles, M. A. Weber, J. Basile, A. H. Gradman, D. B. Bharucha, W. Chen, M. Pattathil, N.-M.-. S. Investigators, et al., Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study, The Lancet 383 (9932) (2014) 1889–1898.

[6] W. Zheng, W. Sun, A. Simeonov, Drug repurposing screens and synergistic drug-combinations for infectious diseases, British journal of pharmacology 175 (2) (2018) 181–191.

[7] Y. Kim, S. Zheng, J. Tang, W. Jim Zheng, Z. Li, X. Jiang, Anticancer drug synergy prediction in understudied tissues using transfer learning, Journal of the American Medical Informatics Association 28 (1) (2021) 42–51.

[8] P. P. Vitiello, G. Martini, L. Mele, E. F. Giunta, V. De Falco, D. Ciardiello, V. Belli, C. Cardone, N. Matrone, L. Poliero, et al., Vulnerability to low-dose combination of irinotecan and niraparib in atm-mutated colorectal cancer, Journal of Experimental & Clinical Cancer Research 40 (1) (2021) 1–15.

[9] T. Liu, R. Yacoub, L. D. Taliaferro-Smith, S.-Y. Sun, T. R. Graham, R. Dolan, C. Lobo, M. Tighiouart, L. Yang, A. Adams, et al., Combinatorial effects of lapatinib and rapamycin in triple-negative breast cancer cells, Molecular cancer therapeutics 10 (8) (2011) 1460–1469.

[10] F. Azam, A. Vazquez, Trends in phase ii trials for cancer therapies, Cancers 13 (2) (2021) 178.

[11] P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, C. Zheng, Z. Guo, X. Chen, W. Zhou, et al., Large-scale exploration and analysis of drug combinations, Bioinformatics 31 (12) (2015) 2007–2016.

[12] R. P. Hertzberg, A. J. Pope, High-throughput screening: new technology for the 21st century, Current opinion in chemical biology 4 (4) (2000) 445–451.

[13] J. Bajorath, Integration of virtual and high-throughput screening, Nature Reviews Drug Discovery 1 (11) (2002) 882–894.

[14] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, et al., Impact of high-throughput screening in biomedical research, Nature reviews Drug discovery 10 (3) (2011) 188–195.

[15] N. P. Torres, A. Y. Lee, G. Giaever, C. Nislow, G. W. Brown, A high-throughput yeast assay identifies synergistic drug combinations, Assay and drug development technologies 11 (5) (2013) 299–307.

[16] D. Ferreira, F. Adega, R. Chaves, The importance of cancer cell lines as in vitro models in cancer methylome analysis and anticancer drugs testing, Oncogenomics and cancer proteomics-novel approaches in biomarkers discovery and therapeutic targets in cancer (2013) 139–166.

[17] K. N. Kragh, D. Gijón, A. Maruri, A. Antonelli, M. Coppi, M. Kolpen, S. Crone, C. Tellapragada, B. Hasan, S. Radmer, et al., Effective antimicrobial combination in vivo treatment predicted with microcalorimetry screening, Journal of Antimicrobial Chemotherapy (2021).

[18] E. S. Potekhina, D. Y. Bass, I. V. Kelmanson, E. S. Fetisova, A. V. Ivanenko, V. V. Belousov, D. S. Bilan, Drug screening with genetically encoded fluorescent sensors: Today and tomorrow, International Journal of Molecular Sciences 22 (1) (2021) 148.

[19] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, et al., The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (7391) (2012) 603–607.

[20] H. Liu, W. Zhang, B. Zou, J. Wang, Y. Deng, L. Deng, Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy, Nucleic acids research 48 (D1) (2020) D871–D881.

[21] J. O'Neil, Y. Benita, I. Feldman, M. Chenard, B. Roberts, Y. Liu, J. Li, A. Kral, S. Lejnine, A. Loboda, et al., An unbiased oncology compound screen to identify novel combination strategies, Molecular cancer therapeutics 15 (6) (2016) 1155–1162.

[22] M. P. Menden, D. Wang, M. J. Mason, B. Szalai, K. C. Bulusu, Y. Guan, T. Yu, J. Kang, M. Jeon, R. Wolfinger, et al., Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen, Nature communications 10 (1) (2019) 1–17.

[23] R. Sałat, K. Sałat, The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy, Computer methods and programs in biomedicine 111 (2) (2013) 330–337.

[24] Y. Qi, Random forest for bioinformatics, in: Ensemble machine learning, Springer, 2012, pp. 307–323.
[25] K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, G. Klambauer, Deepsynergy: predicting anti-cancer drug synergy with deep learning, Bioinformatics 34 (9) (2018) 1538–1546.
[26] Y. Hu, C.-h. Chen, Y.-y. Ding, X. Wen, B. Wang, L. Gao, K. Tan, Optimal control nodes in disease-perturbed networks as targets for combination therapy, Nature communications 10 (1) (2019) 1–14.
[27] J. C. Van Hasselt, R. Iyengar, Systems pharmacology: defining the interactions of drug combinations, Annual review of pharmacology and toxicology 59 (2019) 21–40.
[28] K. E. Regan-Fendt, J. Xu, M. DiVincenzo, M. C. Duggan, R. Shakya, R. Na, W. E. Carson, P. R. Payne, F. Li, Synergy from gene expression and network mining (syngenet) method predicts synergistic drug combinations for diverse melanoma genomic subtypes, NPJ systems biology and applications 5 (1) (2019) 1–15.
[29] F. Cheng, I. A. Kovács, A.-L. Barabási, Network-based prediction of drug combinations, Nature communications 10 (1) (2019) 1–11.
[30] S. Liu, B. Tang, Q. Chen, X. Wang, Drug-drug interaction extraction via convolutional neural networks, Computational and mathematical methods in medicine 2016 (2016).
[31] Z. Sun, S. Huang, P. Jiang, P. Hu, Dtf: Deep tensor factorization for predicting anticancer drug synergy, Bioinformatics 36 (16) (2020) 4483–4489.
[32] S. Loewe, The problem of synergism and antagonism of combined drugs, Arzneimittelforschung 3 (1953) 285–290.
[33] G. Y. Di Veroli, C. Fornari, D. Wang, S. Mollard, J. L. Bramhall, F. M. Richards, D. I. Jodrell, Combenefit: an interactive platform for the analysis and visualization of drug combinations, Bioinformatics 32 (18) (2016) 2866–2868.
[34] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., Drugbank 5.0: a major update to the drugbank database for 2018, Nucleic acids research 46 (D1) (2018) D1074–D1082.
[35] D. Weininger, Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules, Journal of chemical information and computer sciences 28 (1) (1988) 31–36.
[36] G. Landrum, et al., Rdkit: Open-source cheminformatics (2006).
[37] B. Ramsundar, P. Eastman, P. Walters, V. Pande, Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more, " O'Reilly Media, Inc.", 2019.
[38] R. Li, S. Wang, F. Zhu, J. Huang, Adaptive graph convolutional neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018.
[39] Y. Shen, C. Feng, Y. Yang, D. Tian, Mining point cloud local structures by kernel correlation and graph pooling, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4548–4557.
[40] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, et al., Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells, Nucleic acids research 41 (D1) (2012) D955–D961.
[41] L. Cheng, L. Li, Systematic quality control analysis of lincs data, CPT: pharmacometrics & systems pharmacology 5 (11) (2016) 588–598.
[42] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, et al., The gencode v7 catalog of human long noncoding rnas: analysis of their gene structure, evolution, and expression, Genome research 22 (9) (2012) 1775–1789.
[43] J. G. Meyer, S. Liu, I. J. Miller, J. J. Coon, A. Gitter, Learning drug functions from chemical structures with convolutional neural networks and random forests, Journal of chemical information and modeling 59 (10) (2019) 4438–4449.
[44] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Frontiers in neurorobotics 7 (2013) 21.
[45] H. Geppert, T. Horváth, T. Gärtner, S. Wrobel, J. Bajorath, Support-vector-machine-based ranking significantly improves the effectiveness of similarity searching using 2d fingerprints and multiple reference compounds, Journal of chemical information and modeling 48 (4) (2008) 742–746.
[46] C. Crafter, J. P. Vincent, E. Tang, P. DUDlEy, N. H. JAmES, T. Klinowska, B. R. DAvIEs, Combining

azd8931, a novel egfr/her2/her3 signalling inhibitor, with azd5363 limits akt inhibitor induced feedback and enhances antitumour efficacy in her2-amplified breast cancer models, International journal of oncology 47 (2) (2015) 446–454.

[47] B. Alagesan, G. Contino, A. R. Guimaraes, R. B. Corcoran, V. Deshpande, G. R. Wojtkiewicz, A. F. Hezel, K.-K. Wong, M. Loda, R. Weissleder, et al., Combined mek and pi3k inhibition in a mouse model of pancreatic cancer, Clinical cancer research 21 (2) (2015) 396–404.

[48] R. Mäkelä, A. Arjonen, A. S. Rahmanto, V. Härmä, J. Lehtiö, T. Kuopio, T. Helleday, O. Sangfelt, J. Kononen, J. K. Rantala, Ex vivo assessment of targeted therapies in a rare metastatic epithelial–myoepithelial carcinoma, Neoplasia 22 (9) (2020) 390–398.

[49] J. A. Beaver, J. P. Gustin, H. Y. Kyung, A. Rajpurohit, M. Thomas, S. F. Gilbert, D. M. Rosen, B. H. Park, J. Lauring, Pik3ca and akt1 mutations have distinct effects on sensitivity to targeted pathway inhibitors in an isogenic luminal breast cancer model system, Clinical Cancer Research 19 (19) (2013) 5413–5422.

[50] H. Modjtahedi, B. C. Cho, M. C. Michel, F. Solca, A comprehensive review of the preclinical efficacy profile of the erbb family blocker afatinib in cancer, Naunyn-Schmiedeberg's archives of pharmacology 387 (6) (2014) 505–521.

[51] M. J. Akhtar, Covid19 inhibitors: a prospective therapeutics, Bioorg. Chem 101 (2020) 104027.

[52] T. C. Pereira, R. T. de Menezes, H. C. de Oliveira, L. D. de Oliveira, L. Scorzoni, In vitro synergistic effects of fluoxetine and paroxetine in combination with amphotericin b against cryptococcus neoformans, Pathogens and Disease (2021).

[53] J. C. Ontong, N. F. Ozioma, S. P. Voravuthikunchai, S. Chusri, Synergistic antibacterial effects of colistin in combination with aminoglycoside, carbapenems, cephalosporins, fluoroquinolones, tetracyclines, fosfomycin, and piperacillin on multidrug resistant klebsiella pneumoniae isolates, Plos one 16 (1) (2021) e0244673.

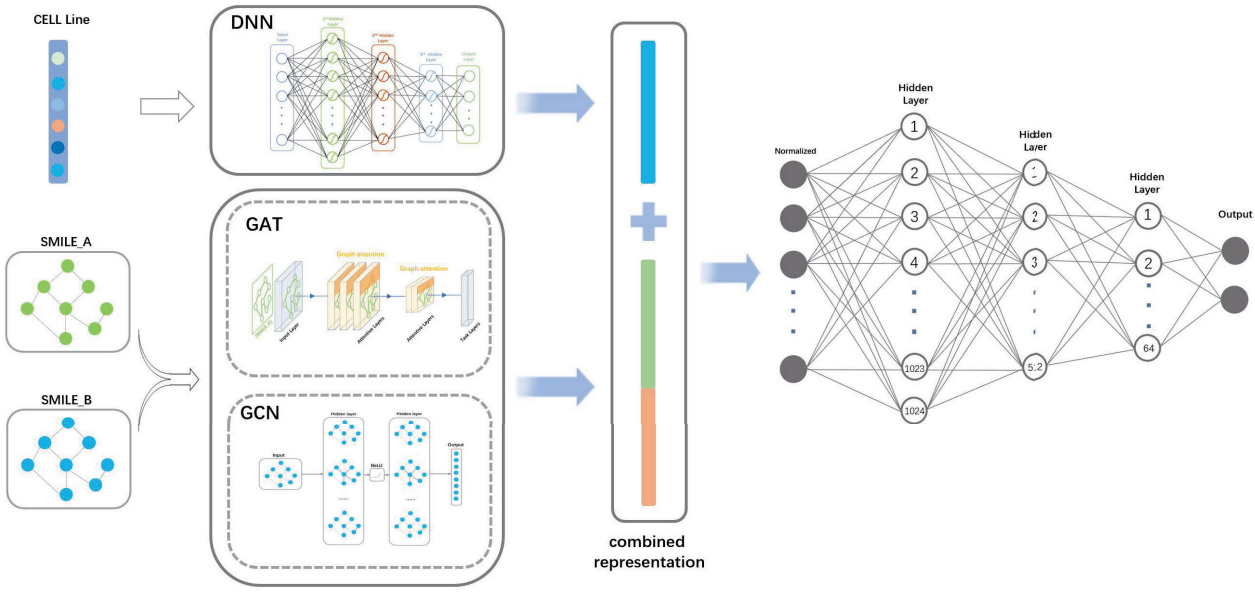

Figure 1: The pipeline of the proposed DeepDDS. The representation of cell lines are obtained through DNN using expression data; the representation of drugs are obtained through GAT and GCN based on the graphs; the embedding of drugs and cell lines are concatenated, which are fed into a fully connected layer to predict the synergistic effect.

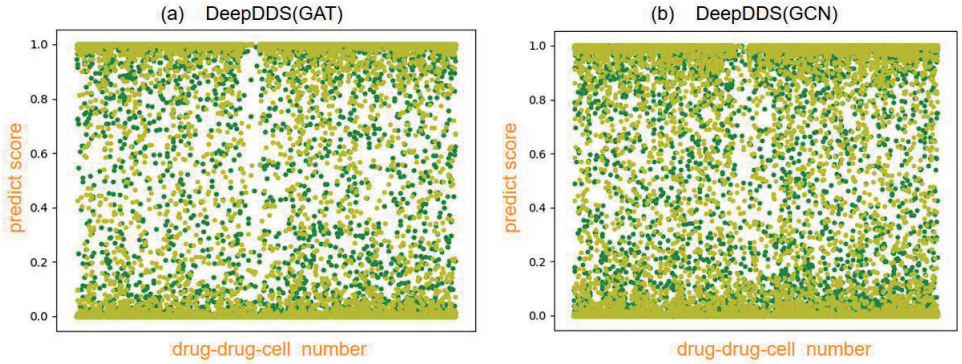

Figure 2: (**a**) In the DeepDDS (GAT) method, the two-class synergy prediction scores of drug pairs combined in different orders, the x-axis is the drug pair, and the y-axis is the synergy score. The yellow dots are the combination of drugA-drugB-cell line, and the green dots are the combination of drugB-drugA-cell line. (**b**) In the DeepDDS (GCN) method, the two-class synergy prediction scores of drug pairs combined in different orders, the x-axis is the drug pair, and the y-axis is the synergy score. The yellow dots are the combination of drugA-drugB-cell line, and the green dots are the combination of drugB-drugA-cell line.

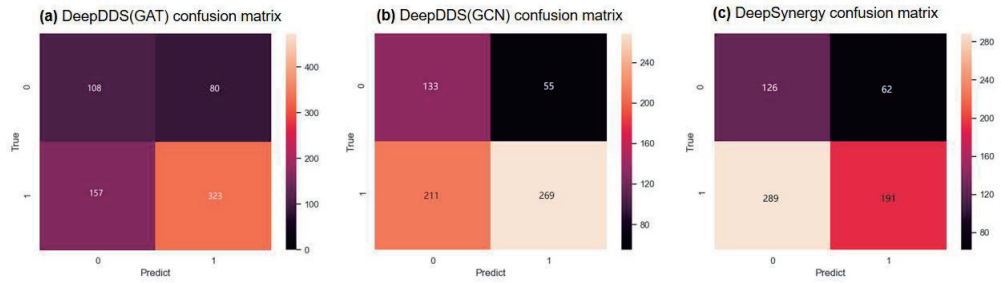

Figure 3: (a) The confusion matrix of DeepDDS (GAT) in the independent test set has an ACC of 0.64. The confusion matrix of DeepDDS (GCN) in the independent test set has an ACC of 0.60. The confusion matrix of DeepSynergy in the independent test set has an ACC of 0.47.

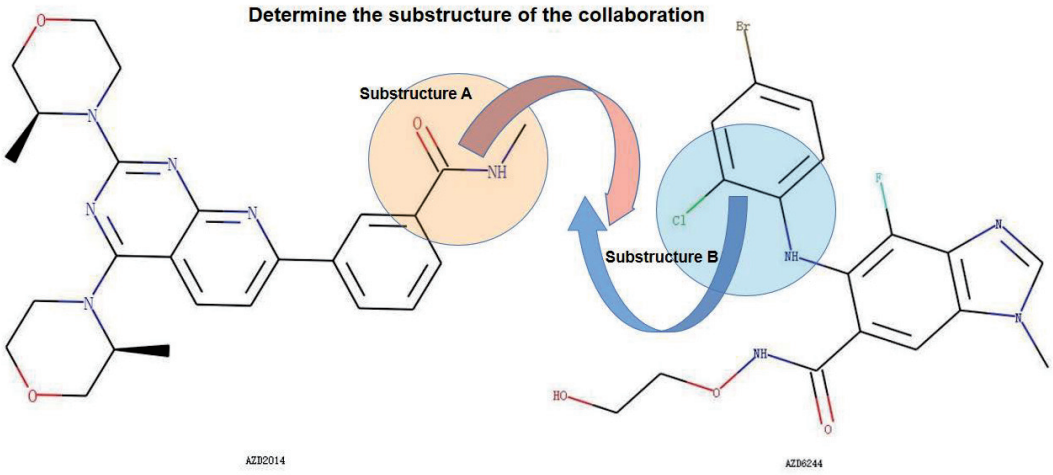

Figure 4: Attention weights in the graph attention network effectively reveal the important substructures of the drugs

Table 1: Performance metrics for the classification task in 5-fold cross-validation

Performance Metric	ROC AUC	PR AUC	ACC	BACC	PREC	TPR	KAPPA
DeepDDS(GAT)	**0.93 ± 0.01**	**0.93 ± 0.01**	**0.85 ± 0.07**	**0.85 ± 0.07**	**0.85 ± 0.07**	**0.85 ± 0.07**	**0.71 ± 0.21**
DeepDDS(GCN)	**0.93 ± 0.01**	**0.92 ± 0.01**	**0.85 ± 0.01**	**0.85 ± 0.01**	**0.85 ± 0.01**	**0.84 ± 0.01**	**0.70 ± 0.22**
DeepSynergy	0.88 ± 0.01	0.87 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.81 ± 0.01	0.75 ± 0.01	0.59 ± 0.15
Random Forests	0.86 ± 0.02	0.85 ± 0.02	0.77 ± 0.01	0.77 ± 0.01	0.78 ± 0.02	0.74 ± 0.01	0.55 ± 0.04
Gradient Boosting Machines	0.85 ± 0.02	0.85 ± 0.01	0.76 ± 0.02	0.76 ± 0.02	0.77 ± 0.01	0.74 ± 0.01	0.53 ± 0.04
Support Vector Machines	0.58 ± 0.01	0.56 ± 0.02	0.54 ± 0.01	0.54 ± 0.01	0.54 ± 0.01	0.51 ± 0.12	0.08 ± 0.04

Table 2: Performance metrics for the classification task in independent test set

Performance Metric	ROC AUC	PR AUC	ACC	BACC	PREC	TPR	KAPPA
DeepDDS(GAT)	0.66 ± 0.12	0.82 ± 0.15	**0.64 ± 0.15**	0.62 ± 0.13	0.80 ± 0.11	**0.67 ± 0.12**	**0.21 ± 0.29**
DeepDDS(GCN)	**0.67 ± 0.12**	**0.83 ± 0.13**	0.60 ± 0.11	**0.63 ± 0.13**	**0.83 ± 0.10**	0.56 ± 0.20	0.21 ± 0.23
DeepSynergy	0.55 ± 0.15	0.71 ± 0.13	0.47 ± 0.14	0.53 ± 0.13	0.75 ± 0.14	0.39 ± 0.17	0.04 ± 0.15
Random Forests	0.53 ± 0.14	0.76 ± 0.16	0.50 ± 0.14	0.54 ± 0.13	0.75 ± 0.14	0.49 ± 0.14	0.06 ± 0.11
Gradient Boosting Machines	0.51 ± 0.10	0.71 ± 0.09	0.45 ± 0.12	0.47 ± 0.08	0.69 ± 0.14	0.43 ± 0.12	-0.03 ± 0.14
Support Vector Machines	0.47 ± 0.11	0.71 ± 0.13	0.54 ± 0.13	0.47 ± 0.15	0.70 ± 0.13	0.63 ± 0.11	-0.04 ± 0.15