Briefings in Bioinformatics, 23(1), 2022, 1–12

https://doi.org/10.1093/bib/bbab390
Problem Solving Protocol

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

Jinxian Wang†, Xuejun Liu†, Siyuan Shen, Lei Deng and Hui Liu

Corresponding author: Lei Deng. Tel/Fax: +86 73182539736; E-mail: leideng@csu.edu.cn; Hui Liu. Tel/Fax: +86 025-58139500; E-mail: hliu@njtech.edu.cn
†These authors contributed equally to this work.

Abstract

Motivation: Drug combination therapy has become an increasingly promising method in the treatment of cancer. However, the number of possible drug combinations is so huge that it is hard to screen synergistic drug combinations through wet-lab experiments. Therefore, computational screening has become an important way to prioritize drug combinations. Graph neural network has recently shown remarkable performance in the prediction of compound–protein interactions, but it has not been applied to the screening of drug combinations.
Results: In this paper, we proposed a deep learning model based on graph neural network and attention mechanism to identify drug combinations that can effectively inhibit the viability of specific cancer cells. The feature embeddings of drug molecule structure and gene expression profiles were taken as input to multilayer feedforward neural network to identify the synergistic drug combinations. We compared DeepDDS (Deep Learning for Drug–Drug Synergy prediction) with classical machine learning methods and other deep learning-based methods on benchmark data set, and the leave-one-out experimental results showed that DeepDDS achieved better performance than competitive methods. Also, on an independent test set released by well-known pharmaceutical enterprise AstraZeneca, DeepDDS was superior to competitive methods by more than 16% predictive precision. Furthermore, we explored the interpretability of the graph attention network and found the correlation matrix of atomic features revealed important chemical substructures of drugs. We believed that DeepDDS is an effective tool that prioritized synergistic drug combinations for further wet-lab experiment validation.
Availability and implementation: Source code and data are available at https://github.com/Sinwang404/DeepDDS/tree/master

Key words: drug combination; attention mechanism; synergistic effect; graph neural network; deep learning; chemical structure.

Jinxian Wang received the Bachelor's degree from Hunan Agricultural University in 2019, and at present is studying for a Master's degree at Central South University supervised by Prof. Lei Deng. His study focuses on machine learning and bioinformatics.
Xuejun Liu is a professor at School of Computer Science and Technology, Nanjing Tech University, Nanjing, China. His research interests include data mining and deep learning.
Siyuan Shen is a graduate student at School of Software, Xinjiang University, Urumqi, China. His research interest is using machine learning algorithms to study noncoding RNA interactions and functions.
Lei Deng Lei Deng is a professor at School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include data mining, bioinformatics and systems biology.
Hui Liu is a professor at School of Computer Science and Technology, Nanjing Tech University, Nanjing, China. His research interests include the anticancer drug screening by means of bioinformatics and deep learning.
Submitted: 7 July 2021; **Received (in revised form):** 14 August 2021

Introduction

Both traditional and modern medicine have taken advantage of the combined use of several active agents to treat diseases. Compared with single-drug therapy, the drug combinations often 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]. For example, triple-negative breast cancer is a malignant tumor with strong invasiveness, high metastasis rate and poor prognosis. Lapatinib or Rapamycin alone has little therapeutic effect, but their combined treatment has been reported to significantly increase the apoptosis rate of triple-negative breast cancer cells [9]. However, some drug combinations may cause antagonistic effects and even aggravate the disease [10]. Therefore, it is crucial to accurately discover synergistic drug combinations to specific diseases.

Traditional discovery of drug combinations is mainly based on clinical trials and limited to only small number of drugs [11], far from meeting the urgent need for anticancer drugs. Due to the great number of possible drug combinations, the traditional method is cost-consuming and impractical. With the development of high-throughput drug screening technology, people can simultaneously carry out large-scale screening of drug combinations over hundreds of cancer cell lines [12–14]. Torres *et al.* [15] utilized yeast to screen a large number of drug combinations and provided a method to identify preferential drug combinations for further testing in human cells. 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 been developed to screen effective antimicrobial combinations for *in vivo* disease treatment. However, these techniques require skilled operations and complicated experimental procedures.

In recent years, the data sets of single drug sensitivities to cancer cell lines increase greatly, such as Cancer Cell Line Encyclopedia (CCLE) [19] and Genomics of Drug Sensitivity in Cancer, which contains drug sensitivities to hundreds of human cancer cell lines, as well as gene expression profiles, mutants and copy number variants. Meanwhile, several large-scale data resources of drug combinations have been released. For example, O'Neil *et al.* [20] released a large-scale drug pair synergy study, which included more than 20 000 pairwise synergy scores between 38 unique drugs. The famous pharmaceutical company AstraZeneca [21] released their drug pair collaboration experiments, which include 11 576 experiments of 910 drug combinations to 85 cancer cell lines with genome-related information. DrugCombDB [22] has collected more than 6 000 000 quantitative drug dose responses, by which they calculated synergy scores to evaluate synergy or antagonism for each drug combination. In addition, quite a few data portals designed to collect drug combinations and relevant knowledge have been developed. The aforementioned data resources motivated the development of computational screening of drug combinations. Many studies have been proposed to explore the vast space of drug combinations to identify synergistic efficacy. For example, classical machine learning methods, such as support vector machine (SVM) and random forest (RF), were successfully predicted the maximal antiallodynic effect of a new derivative of dihydrofuran-2-one (LPP1) used in combination with pregabalin in the streptozocin-induced neuropathic pain model in mice [23, 24]. Liu *et al.* [25] trained a gradient tree boosting classifier to predict new drug combinations using the features by running random walk with restart on the drug–protein heterogeneous network.

Recently, deep learning is increasingly applied to drug development and discovery. For example, DeepSynergy [26] combined the chemical information of drugs and genomic features of cancer cells to predict drug pairs with synergistic effects. TranSynergy [27] is a mechanism-driven and self-attention boosted deep learning model that integrates information from gene–gene interaction networks, gene dependencies, and drug–target associations to predict synergistic drug combinations and deconvolute the cellular mechanisms. MatchMaker [28] takes as input the compound features calculated by ChemoPy [29] and the expression profile of landmark genes [30] together to predict synergistic drug combinations. comboLTR [31] is a new polynomial regression-based framework for modeling anticancer effects of drug combinations in various doses. More interesting, Deng *et al.* [32] presented a pathway-guided deep neural network (DNN) model, which reshapes the model by incorporating a layer of pathway nodes and their connections to input gene nodes, to predict the drug sensitivity in cancer cells. On the other hand, some studies applied Simplified Molecular Input Line Entry System (SMILES) to characterize the chemical properties of drugs. For example, Liu *et al.* regarded the SMILES code as a string and directly input into a convolutional neural network [33] to extract drug features for subsequent prediction task. In addition, graph neural network (GNN) is used to learn feature representation from drug chemical structure [34, 35].

In this paper, we proposed a deep learning model, DeepDDS (Deep Learning for Drug–Drug Synergy prediction), to predict the synergistic effect of drug combinations. At first, the drug chemical structure is represented by a graph in which the vertices are atoms and the edges are chemical bonds. Next, a graph convolutional network (GCN) and attention mechanism is used to compute the drug embedding vectors. By integration of the genomic and pharmaceutical features, DeepDDS can capture important information from drug chemical structure and gene expression patterns to identify synergistic drug combinations to specific cancer cell lines. We compare DeepDDS to both classical machine learning methods [SVM, RF, Adaboost, Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost)] and other latest deep learning (DTF [36], AuDNNsynergy [37], DeepSynergy [26] and TranSynergy [27]) on benchmark data set, DeepDDS significantly outperformed other competitive methods. In particular, we conducted leave-one-out experiments to verify that DeepDDS achieved better performance when one drug (combination) or one tissue is not included in the training set. Also, on an independent test set released by well-known pharmaceutical enterprise AstraZeneca, DeepDDS was superior to competitive methods by more than 16% predictive precision. We also explored the function of graph attention network (GAT) in revealing important chemical substructures of drugs and found the correlation matrix of atomic features showed clustering patterns among atom subgroups during the training process. Finally, we use the trained model to predict novel drug combinations and find five previously reported synergistic drug combinations in the top 10 predicted results. In summary, we believed that DeepDDS is an effective tool that prioritized synergistic drug combinations for further wet-lab experiment validation.

Materials and methods

Data source

The SMILES [38] of drugs are obtained from DrugBank [39], based on which the chemical structure of a drug can be converted to a graph using RDKit [40]. In the molecular graph, the vertices are atoms and the edges are chemical bonds.

The gene expression data of cancer cell lines are obtained from CCLE [19], which is an independent project that makes effort to characterize genomes, messenger RNA expression and anticancer drug dose responses across cancer cell lines. The expression data is normalized through Transcripts Per Million based on the genome-wide read counts matrix.

To construct the benchmark set, we obtain the drug combination sensitivity data from a recently released large-scale oncology screening data set [20], where the viability of 39 cancer cells treated with thousands of drug combinations was evaluated by biochemical assay. The Loewe Additivity score [41], a quantitative metric that defines the synergistic or antagonistic effect of the drug combination, was calculated based on the 4 by 4 dose–response matrix using the Combenefit tool [42]. Of note, multiple replicates of one drug combination were assayed in the original data, and thus the average score of the replicates was selected as the final synergistic score for each unique drug pair–cellline combination. According to the Loewe score, a combination with the score above 0 is regarded as synergistic, and with the score below 0 is antagonistic. Obviously, the drug combinations with higher synergistic scores are more attractive candidates for further clinical experiments. Since many additive combinations may exist (synergy scores are around 0 due to noise), we choose a stricter threshold to classify the combinations. Particularly, combinations with synergy score higher than 10 are labeled as positive (synergistic), and those with score less than 0 are labeled as negative (antagonistic). This yielded a balanced benchmark set that contains 12 415 unique drug pair–cell line combinations, covering 36 anticancer drugs and 31 human cancer cell lines.

Pipeline of DeepDDS

Figure 1 illustrates the end-to-end learning framework for the prediction of synergistic drug combinations. For each pairwise drug combination, the input layer first receives the molecular graphs of two drugs and gene expression profiles of one cancer cell line that was treated by these two drugs. We tested two types of GNNs, GAT and GCN, to extract features of drugs. The genomic feature representation of cancer cells is encoded by a multilayer perception (MLP). The embedding vectors are subsequently concatenated as the final feature representation of each drug paircell line combination, which is propagated through the fully connected layers for the binary classification of drug combinations (synergistic or antagonistic).

Drug representation based on GNN

We use the open-source chemical informatics software RDKit [40] to convert the SMILES into molecular graphs, where the nodes are atoms and the edges are chemical bonds. Specifically, a graph for a given drug is defined as $G = (V, E)$, where V is the set of N nodes that is represented by a C-dimensional vector, and E is the set of edges represented as an adjacency matrix A. In a molecule graph, $x_i \in V$ is the ith atom and $e_{ij} \in E$ is the chemical bond between the ith and jth atoms. The chemical molecular graph is non-Euclidean data and lacks of translation invariance; therefore, we applied GNN instead of traditional convolution network to extract drug feature representations based on the graphs.

For each node in a graph, we use DeepChem [43] to compute a set of atomic attributes as its initial feature. Specifically, each node is represented as a binary vector including five pieces of information: the atom symbol, the number of adjacent atoms, the number of adjacent hydrogen, the implicit value of the atom and whether the atom is in an aromatic structure. In GNN, the learning process of drug representation is actually the message passing between each node and its neighbor nodes. In this paper, we consider two types of GNN (GCN and GAT) in our learning framework and evaluate their performance in the drug feature extraction.

GCN

The input of the multilayer GCN is the node feature matrix $X \in \mathbb{R}^{N\times C}$ and the adjacency matrix $A \in \mathbb{R}^{N\times N}$ that represents the connection of nodes. According to Welling *et al.* [44], it can write dissemination rules in a standardized format to ensure stability.

The iteration process can be defined as below:

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

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 a learnable parameter.

The output $Z \in \mathbb{R}^{N\times F}$ (F is the number of output features per node) can be defined as below:

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

where $\Theta \in \mathbb{R}^{C\times F}$ (F is the number of filters or feature maps) is the matrix of filter parameters.

Our GCN-based model uses three GCN layers activated by rectified linear unit (ReLU) function. The original GCN is a method for classifying the node by semi-supervised learning, i.e. its outputs are the node-level feature vectors. To construct graph-level feature vectors, we use Sum, Average and Max Pooling to aggregate the whole graph feature from learned node features and evaluate their performance. We find that the use of Max Pooling layer in GCN-based DeepDDS outperforms the others. Therefore, we add a global Max Pooling layer after the last GCN layer to extract the representation.

GAT

The GAT proposes a multihead attention-based architecture to learn higher-level features of nodes in a graph by applying a self-attention mechanism. Every attention head has its own parameters. The GAT architecture is built from the graphics attention layer. The output features for nodes were computed as

$$h_i' = ||_{m=1,\ldots,M}\left(\alpha_{i,i}^m W h_i + \sum_{j\in N(i)} \alpha_{i,j}^m W h_j\right) \tag{3}$$

where $\|$ concat the output results of multiple attention mechanisms, M is the number of attention heads and $W \in \mathbb{R}^{C'\times C}$ is a weight matrix. The attention coefficient $\alpha_{i,j}$, between each input

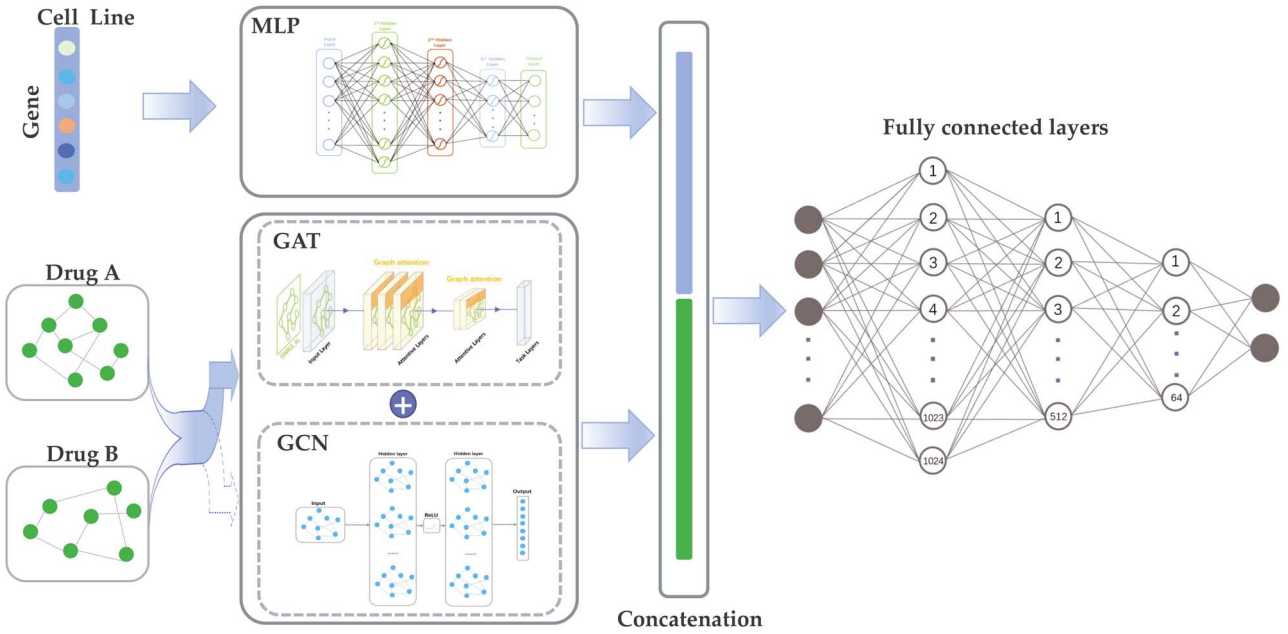

Figure 1. The pipeline of DeepDDS learning framework. The feature embedding of gene expression profiles of cancer cell line is obtained through multilayer perception (MLP), and the feature embedding of drug is obtained through GAT or GCN based on the drug molecular graph generated from drug SMILES. The embedding vectors of drug and cell line are subsequently concatenated to feed into a multilayer fully connected network to predict the synergistic effect.

node i and its 1st-order neighbor in the graph, is calculated as follows:

$$\alpha_{i,j} = \frac{exp(elu(a^T[Wh_i||Wh_j]))}{\sum_{k\in N(i)} exp(elu(a^T[Wh_i||Wh_k]))} \quad (4)$$

where $a^T \in \mathbb{R}^{C'}$ is learnable weight vector, T is the corresponding transpose and elu is a nonlinear activation function, when x is negative, y is equal to 0. Then, 'softmax' function is introduced to normalize all neighbor nodes j of i for easy calculation and comparison.

Cell line feature extraction based on MLP

To alleviate the dimension imbalance between the feature vectors of drugs and cell lines, we selected the significant genes according to a The Library of Integrated Network-Based Cellular Signatures (LINCS) project [30]. 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 based on the Connectivity Map data [45]. The intersected genes between the CCLE gene expression profiles and the landmark set were chosen for subsequent analysis. We used the gene annotation information in the CCLE[19] and the GENCODE annotation database [46] to remove the redundant data, as well as the transcripts of noncoding RNA. Finally, we select **954** genes from raw expression profiles as input to the model.

We adopt an MLP to extract the cell line features. The MLP includes two hidden layers, and the number of hidden units of each layer is selected via hyperparameter selection (see Hyperparameter setting for detail).

Predicting the synergistic effect of drug combinations versus cell lines

We formulated the prediction of synergistic drug combinations as an end-to-end binary classification model. Upon the embedding vectors of drugs through GAT or GCN, and the embedding vectors of cell lines through MLP, they are concatenated as the input of multiple fully connected layers. We adopt the spindle-shaped structure for the fully connected layer. The probability of the synergistic effect (classification label) was computed by the softmax function that follows the output of the last hidden layer, as follows:

$$p_t = \text{softmax}\left(W_{out} \cdot a^l + b_{out}\right) \quad (5)$$

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:

$$a^l = \sigma(W^l a^{l-1} + b^l) \quad (6)$$

where l is the number of hidden layers, W and b are the matrices corresponding to all hidden layers and output layers, bias vector, $a^0 = \text{concat}(R_{drug1}, R_{drug2}, R_{cellline})$ is the raw input vector.

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

$$F = \min\left(-\sum_{i=1}^{N} \log P_{t_i} + \frac{2}{\lambda}\|\Theta\|\right) \quad (7)$$

where Θ represents the set of all trainable weight and bias parameters involved in the model, N is the total number of samples in the training data set, t_i is the ith label and λ is an L2 regularization hyperparameter.

Table 1. Hyperparameter settings of DeepDDS

Hyperparameter	Values
GCN hidden units	[1024,156];**[1024,512,156]**;[512,256,156]
GAN hidden units	**[1024,512]**;[512,128];[1024,156];[512,156]
GAN attention Head	4, 8, **10**, 12, 16
MLP hidden units	[1024,512];**[2048,512]**;[2048,1024];[4096,512]
FC hidden units	[4096, 1024, 512]; [2048, 1024, 512]; **[1024, 512, 128]**; [1024, 512, 64]
Learning rate	10^{-2};**10^{-3}**; 10^{-4}; 10^{-5}
Dropout	No dropout,0.1; **0.2**; 0.3; 0.4; 0.5

The bold values represent the optimal parameter values in our model.

Result

Hyperparameter setting

The real architecture of DeepDDS is actually determined by hyperparameter setting. The hyperparameters cover the numbers of layers and units of each layer in MLP, GCN and GAN, as well as the activation function and learning rate. As exhaustive enumerations of the hyperparameters are computationally inhibitive, thereby we adopt a grid-like search strategy to tune the hyperparameters. As shown in Table 1, we have tested different structural forms and values of these hyperparameters. We tuned the hyperparameters via 5-fold cross-validations on benchmark data set. The selected values of these hyperparameters are displayed in boldface. The GCN yielded better performance in the drug feature extraction when its structure has three hidden layers and the number of units are 1024, 512 and 156, respectively. We have also considered different number of hidden layers for GAN and MLP and found they performed best with two hidden layers. For multihead attention mechanism, multiple independent values are evaluated. For the activation function, the exponential linear unit (ELU) and ReLU activation functions after the GAT layers at DeepDDS-GAT are used. For DeepDDS-GCN, it also has similar network structure, but only ReLU is used as an activation function.

Performance comparison on cross-validation

To evaluate the performance of DeepDDS, we compared DeepDDS with some current state-of-the-art methods, including both classical machine learning methods and deep learning-based methods. Six classical machine learning methods, including RF, GBM, XGBoost, Adaboost, MLP, SVMs, are considered in the performance comparison. Four deep learning-based methods are TranSynergy [27], AuDNNsynergy [37], DeepSynergy [26] and Deep Tensor Factorization (DTF) [36]. To clarify the difference between DeepDDS and these deep learning-based methods, we summarize them as below:

- **TranSynergy:** TranSynergy includes three major components: input dimension reduction component, self-attention transformer component and output fully connected component. It combines the network propagated drug target profile, gene dependency and gene expression to find novel genes associated with the synergistic drug combination from the learned biological relations.
  em **AuDNNsynergy:** AuDNNsynergy integrated multitype of genomic data from The Cancer Genome Atlas database [47], and then utilize transfer learning to improved the prediction accuracy.
- **DeepSynergy:** DeepSynergy uses molecular chemistry and cell line genomic information as input, and a cone layer in a neural network (DNN) to simulate drug synergy and finally predict the synergy score.
- **DTF:** DTF combine tensor-based framework and deep learning methods together to predict synergistic effect of drug pairs, which is comprised mainly of a tensor factorization method and a DNN.

First, we conducted 5-fold cross-validation to benchmark the predictive power of DeepDDS. The training samples (each sample is a drug–drug–cell line triplet) are randomly split into five subsets of roughly equal size, each subset is taken in turn as a test set and the remaining four subsets are used to train the model, whose prediction accuracy on the test set is then evaluated. The average prediction accuracy over the 5-fold is used as the final performance measure. For clarity, we provide typical performance measures widely used in 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 Kappa. Table 2 shows these performance measures of DeepDDS and other methods. Note that if we failed to run one method (after we have tried best) or its source code is unavailable, we used the performance metrics reported in the original context or did not display it in our experimental results. Clearly, DeepDDS-GAT achieved higher accuracy than all other methods, and its performance measures of ROC AUC, PR AUC, ACC, BACC, PREC, TPR, TNR and Kappa reach 0.93, 0.93, 0.85, 0.85, 0.85, 0.85, 0.85 and 0.71, respectively. In fact, both DeepDDS-GAT and DeepDDS-GCN outperform others in terms of all these performance measures. We note that the classifier XGBoost also achieved remarkable performance, nevertheless still inferior to DeepDDS. The three deep learning-based methods, TranSynergy, DTF and DeepSynergy, follow closely XGBoost, but outperform other methods.

We further checked the top 100 drug pairs with the highest predicted synergy scores by DeepDDS-GAT (for detail, see Supplementary Table S1) and found that 98 drug pairs have been experimentally validated to be synergistic combinations over different cancer cell lines.

Performance evaluation by input permutation

We found that the higher the real synergy score, the higher the predictive score. After normalization of real synergy scores to [0, 1] region, we draw a scatter plot of the drug combinations with respect to the predicted and real synergy scores. As shown in Figure 2(a), most points were located closely to the identity line. The Pearson correlation between the predicted synergy scores and real synergy scores reach 0.801. We reported the performance metrics like Mean Absolute Error, Root Mean Square Error and Pearson correction, as shown in Supplementary Table S7. We found that DeepDDS outperformed all other regression models, followed by DeepSynergy, whereas SVR performance worst. The results indicate our method achieve superior predictive accuracy.

Moreover, we verified the predictive performance of DeepDDS upon different input orders of two drugs. For drug A and drug B, we permutate the input features so that drug A–drug B and drug B–drug A are regarded as two different samples to train the model. Figure 2(b) shows the predicted synergy score upon different sequence of input features by DeepDDS-GAT. It can be found that most values locate closely to the identity line and the Pearson correlation coefficient reach 0.9. It can prove that our model is insensitive to the sequence of the input features of drug combinations. In addition, we found that the ROC AUC and PR

Table 2. Performance comparison of DeepDDS and competitive methods on 5-fold cross-validation

Method	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**
XGBoost	0.92 ± 0.01	0.92 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.68 ± 0.01
Random Forest	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
GBM	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
Adaboost	0.83 ± 0.01	0.83 ± 0.03	0.74 ± 0.01	0.74 ± 0.02	0.74 ± 0.02	0.72 ± 0.01	0.48 ± 0.03
MLP	0.65 ± 0.02	0.63 ± 0.05	0.56 ± 0.06	0.56 ± 0.05	0.54 ± 0.04	0.53 ± 0.22	0.12 ± 0.04
SVM	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
AuDNNsynergy	0.91 ± 0.02	0.63 ± 0.06	**0.93 ± 0.01**	NA[a]	0.72 ± 0.06	NA	0.51 ± 0.04
TranSynergy	0.90 ± 0.01	0.89 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.80 ± 0.01	0.64 ± 0.01
DTF	0.89 ± 0.01	0.88 ± 0.01	0.81 ± 0.01	0.81 ± 0.01	0.82 ± 0.01	0.77 ± 0.03	0.63 ± 0.04
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 ± 0.05

[a]NA means we failed to run the source code of corresponding method.
The bold values represent the optimal performance over all competitive methods.

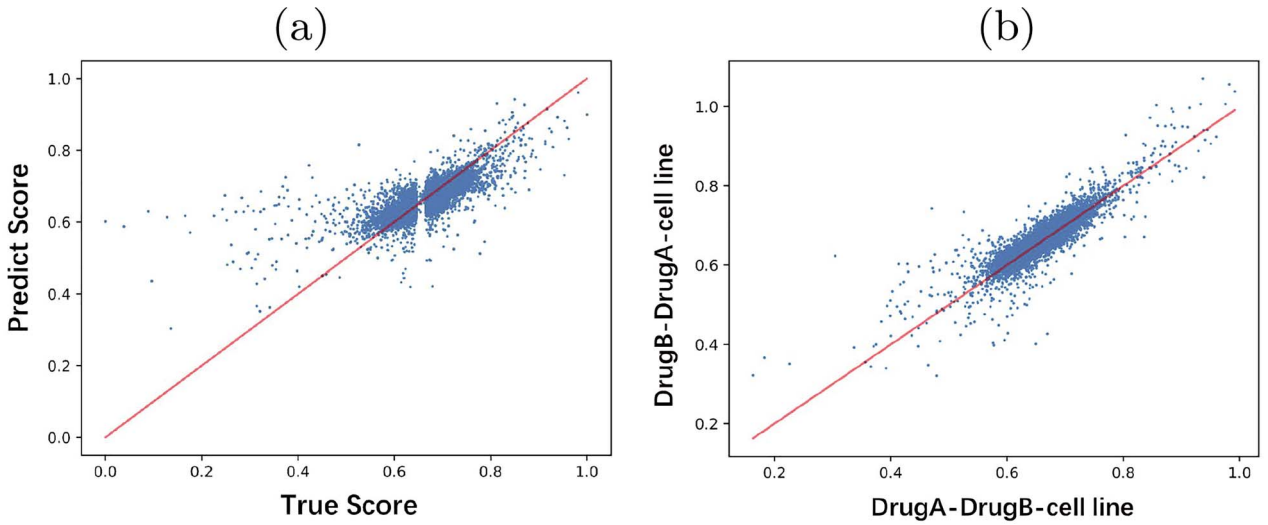

Figure 2. Scatter plots of synergy scores. (**a**) The scatter plot with respect to the real synergy scores and predicted synergy score. (**b**) The scatter plot of synergy score obtained from different input order of two drugs.

AUC obtained by drug A–drug B and drug B–drug A both reach or be close to 0.93.

Performance evaluation by leave-one-out cross-validation

We went further to verify the performance of the DeepDDS model using leave-one-out cross-validations. First, we conducted the leave-one drug combination-out experiment. More precisely, we iteratively excluded each drug combination from the training set and used the remaining data to train the DeepDDS model that was used to predict the sensitivity of the excluded drug combination to cancer cell lines. The result of the leave-one drug combination-out experiment is shown in Table 3, DeepDDS-GAT achieved notably performance by AUC value 0.89, followed by DeepDDS-GCN. It can also be found that DeepDDS significantly outperformed all other methods.

As the leave-one drug combination-out experiment did not exclude single drug entirely from the training set, we next leave one drug out to prevent the information of certain drug being seen by the model. The leave-one drug-out experiment can check the capacity to learn the important features of unseen drugs from the chemical structures of those seen drugs. As shown in Table 3, DeepDDS still achieved better performance than other competitive methods.

As previous studies [27], we also carried out leave-one cell line-out experiment to verify the performance of DeepDDS. Take the cell line T47D as an example, the drug combination between BEZ-235 and MK-8669, Dasatinib, Lapatinib, Geldanamycin, PD325901, Erlotinib, MK-4541, Temozolomide, Vinorelbine, ABT-888, all have high experimental synergy scores (Loewe>100). Expectedly, the prediction scores of these drug combinations have prior rankings among all candidate drug pairs (see Supplementary Tables S2 and S3 for detail). In addition to the leave-one cell line-out evaluation [26], we adopted a more rigorous strategy to evaluate our method. We excluded all the cancer cell lines that belong to specific tissue from the training set, so that the model can not see any gene expression information of a certain type of tissue. We iteratively used the excluded cancer cell lines as the validation set and the remaining samples as the training set to train the model. Table 3 illustrates that DeepDDS-GAT achieved the best performance on leave-one tissue-out evaluation. Also, DeepDDS performs better than all classical machine learning methods and deep learning-based methods. Moreover, Figure 3 shows the ROC AUC values of DeepDDS-GAT, DeepSynergy and TranSynergy

Table 3. Performance on DeepDDS and competitive methods on leave-drug combination-out, leave-drug-out and leave-tissue-out experiments

Method	Leave-drug combination-out			Leave-drug-out			Leave-tissue-out		
	ROC AUC	PR AUC	ACC	ROC AUC	PR AUC	ACC	ROC AUC	PR AUC	ACC
DeepDDS-GAT	**0.89 ± 0.02**	**0.88 ± 0.06**	**0.81 ± 0.03**	**0.73 ± 0.01**	**0.72 ± 0.05**	**0.66 ± 0.02**	**0.83 ± 0.04**	**0.82 ± 0.4**	**0.74 ± 0.03**
XGBoost	0.84 ± 0.02	0.83 ± 0.04	0.75 ± 0.02	0.66 ± 0.09	0.65 ± 0.06	0.61 ± 0.06	0.82 ± 0.01	0.81 ± 0.01	0.73 ± 0.01
TranSynergy	NA[a]	NA	NA	NA	NA	NA	0.81 ± 0.01	0.79 ± 0.02	0.73 ± 0.03
DeepSynergy	0.83 ± 0.03	0.81 ± 0.05	0.77 ± 0.03	0.71 ± 0.07	0.64 ± 0.06	0.61 ± 0.07	0.80 ± 0.01	0.79 ± 0.04	0.71 ± 0.05
Random Forest	0.82 ± 0.02	0.81 ± 0.03	0.74 ± 0.02	0.67 ± 0.08	0.66 ± 0.05	0.62 ± 0.06	0.80 ± 0.08	0.80 ± 0.05	0.71 ± 0.05
MLP	0.82 ± 0.03	0.81 ± 0.05	0.74 ± 0.02	0.69 ± 0.05	0.68 ± 0.04	0.62± 0.06	0.77 ± 0.07	0.76 ± 0.05	0.70 ± 0.06
GBM	0.81 ± 0.03	0.81 ± 0.04	0.74 ± 0.02	0.64 ± 0.09	0.63 ± 0.09	0.60 ± 0.06	0.81 ± 0.08	0.81 ± 0.05	0.72 ± 0.06
Adaboost	0.77 ± 0.02	0.78 ± 0.02	0.69 ± 0.03	0.62 ± 0.11	0.61 ± 0.06	0.58 ± 0.11	0.77 ± 0.12	0.78 ± 0.11	0.70 ± 0.11
SVM	0.66 ± 0.01	0.65 ± 0.05	0.58 ± 0.01	0.60 ± 0.02	0.59 ± 0.05	0.55 ± 0.03	0.66 ± 0.04	0.66 ± 0.07	0.59 ± 0.05

[a]NA means we failed to run the source code of corresponding method.
The bold values represent the optimal performance over all competitive methods.

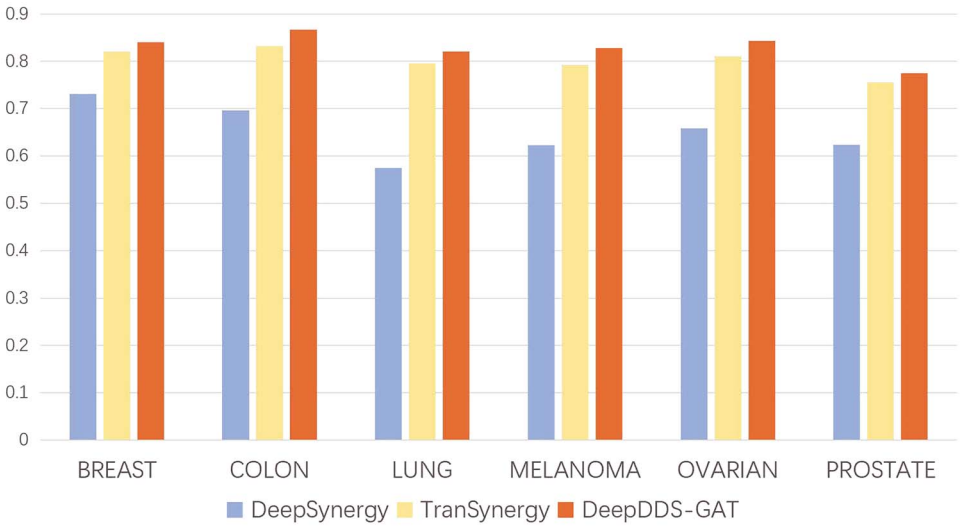

Figure 3. The ROC AUC values of DeepDDS-GAT, DeepSynergy and TranSynergy upon leave-tissue-out evaluations on six different tissues, including breast, colon, lung, melanoma, ovarian and prostate.

on six different tissues, including breast, colon, lung, melanoma, ovarian and prostate. It can be found that DeepDDS-GAT is better than the other two deep learning-based methods with ROC AUC 0.84, 0.867, 0.821, 0.828, 0.843 and 0.775 by leave-one tissue-out cross-validation, respectively.

Evaluation on independent test set

To verify the generalization ability of our method, we use the benchmark data set [20] to train our model, and then employ an independent test set released by AstraZeneca [21] to evaluate the performance of DeepDDS and other competitive methods. The independent test set contains 668 unique drug pair–cell line combinations, covering 57 drugs (Supplementary Table S4) and 24 cell lines (Supplementary Table S5).

Table 4 shows the performance achieved by DeepDDS and competitive methods on the independent test set. It can be seen that the performance of DeepDDS is better than all competitive methods in terms of every performance measure. For clarity, we draw the ROC curves of DeepDDS and other methods, as shown in Figure 4. DeepDDS-GAT and DeepDDS-GCN account for the top two, followed by DeepSynergy. Meanwhile, it can be found that most machine learning-based methods perform just as random guess. This result indicated that classical machine learning methods ran into overfitting, whereas deep learning-based methods acquired better generalization abilities. In particular, DeepDDS-GAT and DeepDDS-GCN correctly predicted 421 (421/668=0.63) and 402 (402/668=0.6) drug pairs included in the independent test set, which outperformed DeepSynergy correct prediction 317 (317/668=0.47) by 16 and 13%, respectively. The confusion matrices in Supplementary Figure S3 show detailed numbers of correctly and falsely predicted samples by the three methods.

GAT reveals important chemical substructure

Our DeepDDS-GAT model iteratively passes messages between nodes so that each node can capture the information of its neighboring nodes. Meanwhile, each neuron is connected to the upper neighborhood layer through a set of learnable weights in the GAT network. As a result, the feature representation actually encodes the information of the chemical substructure around the atom, including formal charge, water solubility and other physicochemical properties. This motivates us to explore the implications of the attention mechanism in revealing the important chemical substructures.

For example, previous studies have shown that epidermal growth factor receptor (EGFR) inhibitor Afatinib and serine/threonine protein kinase B (AKT) inhibitor MK2206 played synergistic

Table 4. 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 Forest	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
MLP	0.53 ± 0.13	0.74 ± 0.12	0.53 ± 0.15	0.53 ± 0.15	0.74 ± 0.13	0.53 ± 0.13	0.05 ± 0.11
GBM	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
XGBoost	0.52 ± 0.11	0.73 ± 0.12	0.45 ± 0.15	0.49 ± 0.11	0.71 ± 0.09	0.38 ± 0.17	-0.01 ± 0.14
Adaboost	0.49 ± 0.09	0.69 ± 0.14	0.46 ± 0.17	0.47 ± 0.12	0.69 ± 0.14	0.46 ± 0.15	-0.05 ± 0.17
SVM	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

The bold values represent the optimal performance over all competitive methods.

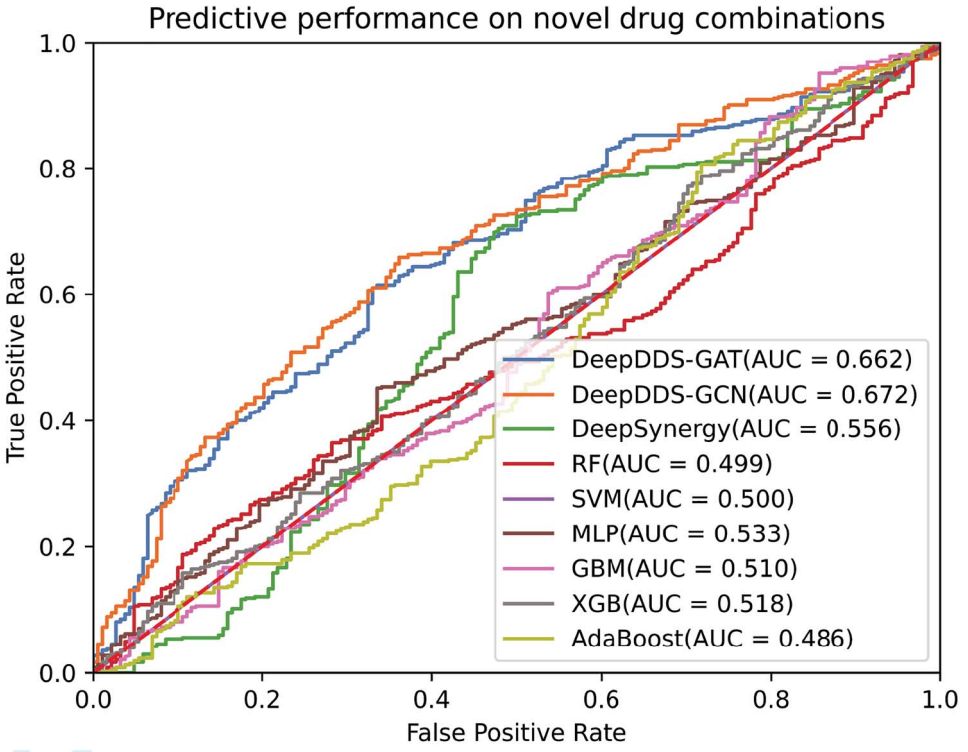

Figure 4. ROC curves and AUC values of DeepDDS and competitive methods on independent test data set released by AstraZeneca.

effects in treating lung cancer, head and neck squamous cell carcinoma [48–50]. We set about to investigate how the atomic feature vectors evolved during the learning process, by measuring the Pearson correlation coefficient between atom pairs based on the feature vectors. The heat maps of the atom correlation matrix were plotted to observe the change of feature patterns. The similarity scores are displayed in the cells and indicated by the color scheme. It can be seen that before training the visual patterns in the heat maps of two drugs shows some degree of chaos. After training, however, the heat map of both drugs show obvious atomic clusters in a specific order. In particular, the atom of drug Afatinib was clustered into five subgroups, whereas MK2206 was clustered into two atom subgroups (one big and one small block), as shown in Figure 5. Without loss of generalization, we randomly selected a few other drug combinations to check whether their feature vectors undergo similar pattern changes during the training process. These drug combinations include AZD2014 and AZD6244, AZD8931 and AZD5363, GDC0941 and AZD6244,GDC0941 and MK2206. As expected, the atomic feature vectors of the involved drugs gradually cluster into several subgroups (see Supplementary Figures S2–S6 for more detail).

We go a step further to explore the interpretability of the GAT in revealing the chemical substructures that are potential components exerting synergistic effect of the drug combinations. We computed the Pearson correlation coefficients between atom pairs across two drugs, so that significant associations between chemical subgroups of different drugs can be uncovered. Take the drug combinations Afatinib and MK2206 as an example again, we found that the heat map of the atom correlation matrix has no clear clustering pattern before training, whereas it shows two notably linking blocks after training, as shown in Figure 6(a). More interesting, these two linking blocks exactly indicate that the bigger atom subgroup (No.1–25 atoms) of MK2206 associates to the 3rd and 5th atom subgroups (No.9–14 atoms and No.21–33 atoms) of Afatinib. From the 3D structures of the two drugs, they are just the main functional groups of Afatinib and MK2206, respectively. For other examples mentioned above, we found that their interdrug atom correlation matrices also

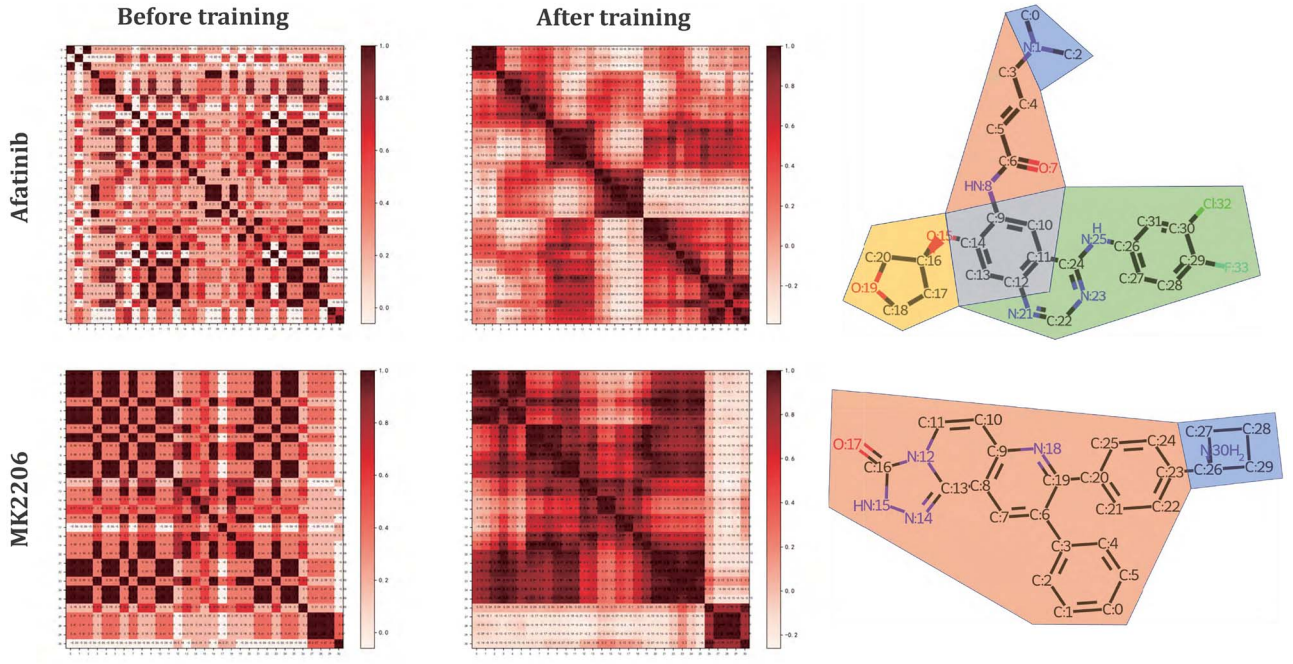

Figure 5. Heat map of the atomic feature similarity matrices of Afatinib and MK2206 before and after training. The heat maps show clear clustering patterns during the learning process. The diagrams of chemical structures of Afatinib and MK2206 display the five and two subgroups according to their clusters of heat maps.

display clustering patterns, as shown in Supplementary Figures S3–S6.

As a result, the atom embedding vectors display clear feature patterns during the training process, namely, the atom correlation matrices clearly cluster into several atom subgroups, and the degree of association between atom subgroups of different drugs transfer from chaos to order. We adventure to speculate that the atom subgroups included in these two drugs play key role in their synergistic function, although the pharmacological mechanism *in vivo* remains unclear to date.

Predicting novel synergistic combinations

The performance evaluation experiments above have shown that our DeepDDS model achieved superior performance, thereby we applied DeepDDS to predict novel synergistic combinations. We used the O'Neil drug combination data set to train the DeepDDS model. To generate candidate drug combinations, we selected 42 small molecule targeted drugs approved by the food and drug administration (FDA) [51] and then generated 855 candidate drug pairs (see Supplementary Table S6). We listed the top 10 predicted drug combinations in Table 5. To verify the reliability of the predicted results, we conducted a nonexhaustive literature search and found there are at least five predicted drug combinations are consistent with the observations in previous studies or under clinical trials. We presented the pubMed unique identifier (PMID)s or digital object identifier (DOI)s of these related publications in Table 5.

For example, the CDK4/6 inhibitor **abemaciclib** and EGFR inhibitor **lapatinib** significantly enhanced growth inhibitory for HER2-positive breast cancer [52]. Ye *et al.* [53] found that **Copanlisib** reduced **Sorafenib**-induced phosphorylation of phospho-Akt (p-AKT) and enhanced synergistically of antineoplastic effect *in vitro*. Also, the combination of **Erlotinib** and **Regorafenib** in the treatment of hepatocellular carcinoma successfully overcome the interference of epidermal growth factors [54]. Addition of **Sorafenib** to **Vemurafenib** increased reactive oxygen species (ROS) production through ferroptosis, thus increasing the sensitivity of melanoma cells to vemurafenib [55]. Zhang *et al.* [56] reproted that the **Regorafenib** combined with the **Lapatinib** could improve antitumor efficacy in human colorectal cancer. We believe that other predicted drug pairs are also promising combinations await for further validation.

Discussion and Conclusion

In this paper, we have proposed a novel method to predict synergistic drug combinations to specific cancer cells. Overall, our method performs significantly better than other competitive methods on the 5-fold cross-validation experiments. However, we noticed that the predictive accuracy of our method is still limited on the independent test set, although the performance of our method is greatly superior than all competitive methods. We think the limited performance is mainly attributed to the small number of training samples. In fact, the benchmark data set actually includes only 36 unique drugs and 31 cancer cell lines, whereas the space for possible drug combinations is much larger when novel drugs are included. We also explored the semantic correlation of the learn node embedding to drug functional groups, whereas previous models like DTF, AuDNNsynergy, DeepSynergy and TranSynergy have poor drug-level interpretability.

We used two different GNNs, GAT and GCN, to learn drug embedding vectors in our method. Overall, GAT performed slightly better than GCN. However, we have realized that the physicochemical properties of the molecular graph and attention weights between the atoms have not been fully exploited. A few recent studies pay increasing attention to knowledge graph embedding. Lin *et al.* [57] proposed an end-to-end framework called Knowledge Graph Neural Network (KGNN) to explore the topological structure of drugs in knowledge graph. Zheng *et al.* introduced PharmKG [58], a dedicated knowledge graph that combined global network structure and heterogeneous domain features. DTiGEMS+ [59] combined graph embedding

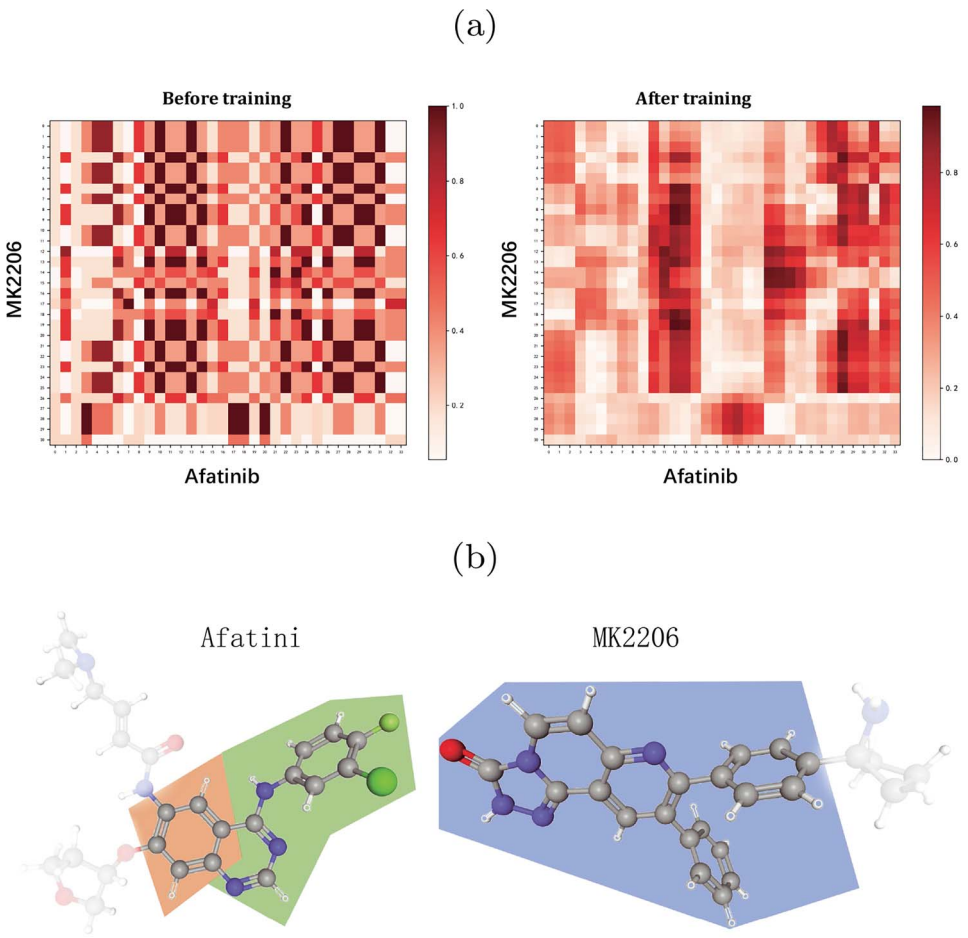

Figure 6. The heat maps of Pearson correlation coefficients between atom pairs across Afatinib and MK2206. Pearson correlation coefficients are computed using the feature vector before and after training. **(a)** The heat map shows no clear visual pattern before training, but after training shows two clear linking blocks. **(b)** The bigger atom subgroup (No.1–25 atoms) of MK2206 associates remarkably to the 3rd and 5th atom subgroups (No.9–14 atoms and No.21–33 atoms) of Afatinib.

Table 5. Top 10 predicted novel synergistic combinations on A375 cancer cell line

Drug A	Drug B	Cell line	Predict score	Publications
Abemaciclib	**Lapatinib**	A375	0.9977	26977878, 33389550, 26977873
Binimetinib	Sorafenib	A375	0.9974	NA
Copanlisib	Regorafenib	A375	0.9973	NA
Copanlisib	**Sorafenib**	A375	0.9973	30962952, 27259258, doi:10.5282/edoc.24304
Binimetinib	Regorafenib	A375	0.9971	NA
Erlotinib	**Regorafenib**	A375	0.997	25907508
Vemurafenib	**Sorafenib**	A375	0.9969	33119140, 30076136, 30844744, 29605720, doi:10.21037/tcr.2020.01.62
Vemurafenib	Regorafenib	A375	0.9967	NA
Lapatinib	**Regorafenib**	A375	0.9967	27864115, 24911215
Pazopanib	Sorafenib	A375	0.9965	NA

The bold values represent the predicted drug combinations that have been previously reported in other papers.

and machine learning to predict drug–target interactions. On the other hand, some studies used self-supervised learning, especially contrastive learning, to learn latent representation from a large number of unlabeled data. Zagidullin *et al.* [60] extracted drug fingerprints by an unsupervised encoder–decoder model and then used them in cancer drug combination discovery. However, the performance of self-supervised learning and transfer learning is still inferior to supervised learning, and their performance decreases significantly in transfer to new tasks.

In conclusion, we have proposed a novel method DeepDDS to predict the synergy of drug combinations for cancer cell lines with high accuracy. Our performance comparison experiments showed that DeepDDS performs better than other competitive methods. We have demonstrated that DeepDDS achieve state-of-the-art performance in a cross-validation setting with an independent test set. We believed that with the increasing size of the data set available, DeepDDS can be further improved and applied to other fields where drug combinations play an essential role,

such as antiviral [61], antifungal [62] and multidrug synergy prediction [63]. Overall, our method yield an inspiring insight into the discovery of synergistic drug combinations.

Key Points

- Combinatorial therapy is a promising method to overcome the resistance of cancer cells to single-target treatments, which motivate us to develop deep learning-based method to predict synergistic drug combinations for specific cancers.
- Two GNN models, GCN and GAT, are employed and compared for their performance in extracting the feature embeddings of drugs.
- We explored the interpretability of the GAT in revealing the important chemical substructures of drugs. Both intra-drug and inter-drug atomic correlations clustered into subgroups, which potentially indicate the pharmacological functional groups governing the synergistic effect of the drug combinations.

Authors' contributions statement

J.W. and H.L. conceived the main idea and the framework of the manuscript. J.W. drafted the manuscript. J.W. and X. L. collected the data and performed the experiments. L.D. and H.L. helped to improve the idea and the manuscript. S.S. reviewed drafts of the paper. L.D. and H.L. supervised the study and provided funding. All authors read and commented on the manuscript.

Funding

The National Natural Science Foundation of China (grants No. 61972422 and No. 62072058).

References

1. Csermely P, Korcsmáros T, Kiss HJM, *et al*. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013; **138**(3): 333–408.
2. Zhao S, Nishimura T, Chen Y, *et al*. Systems pharmacology of adverse event mitigation by drug combinations. *Sci Transl Med* 2013; **5**(206): 206ra140–0.
3. Hill JA, Ammar R, Torti D, *et al*. Genetic and genomic architecture of the evolution of resistance to antifungal drug combinations. *PLoS Genet* 2013; **9**(4):e1003390.
4. Verderosa AD, Dhouib R, Hong Y, *et al*. A high-throughput cell-based assay pipeline for the preclinical development of bacterial dsba inhibitors as antivirulence therapeutics. *Sci Rep* **11**(1): 1–13.
5. Giles TD, Weber MA, Basile J, *et al*. NAC-MD-01 Study Investigators, *et al*. Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *The Lancet* 2014; **383**(9932): 1889–98.
6. Zheng W, Sun W, Simeonov A. Drug repurposing screens and synergistic drug-combinations for infectious diseases. *Br J Pharmacol* 2018; **175**(2): 181–91.
7. Kim Y, Zheng S, Tang J, *et al*. Anticancer drug synergy prediction in understudied tissues using transfer learning. *J Am Med Inform Assoc* 2021; **28**(1): 42–51.
8. Vitiello PP, Martini G, Mele L, *et al*. Vulnerability to low-dose combination of irinotecan and niraparib in atm-mutated colorectal cancer. *J Exp Clin Cancer Res* 2021; **40**(1): 1–15.
9. Liu T, Yacoub R, Taliaferro-Smith LTD, *et al*. Combinatorial effects of lapatinib and rapamycin in triple-negative breast cancer cells. *Mol Cancer Ther* 2011; **10**(8): 1460–9.
10. Azam F, Vazquez A. Trends in phase ii trials for cancer therapies. *Cancer* 2021; **13**(2): 178.
11. Li P, Huang C, Yingxue F, *et al*. Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015; **31**(12): 2007–16.
12. Hertzberg RP, Pope AJ. High-throughput screening: new technology for the 21st century. *Curr Opin Chem Biol* 2000; **4**(4): 445–51.
13. Bajorath J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* 2002; **1**(11): 882–94.
14. Macarron R, Banks MN, Bojanic D, *et al*. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 2011; **10**(3): 188–95.
15. Torres NP, Lee AY, Giaever G, *et al*. A high-throughput yeast assay identifies synergistic drug combinations. *Assay Drug Dev Technol* 2013; **11**(5): 299–307.
16. Ferreira D, Adega F, Chaves R. 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;139–66.
17. Kragh KN, Gijón D, Maruri A, *et al*. Effective antimicrobial combination in vivo treatment predicted with microcalorimetry screening. *Journal of Antimicrobial Chemotherapy* 2021.
18. Potekhina ES, Bass DY, Kelmanson IV, *et al*. Drug screening with genetically encoded fluorescent sensors: Today and tomorrow. *Int J Mol Sci* 2021; **22**(1): 148.
19. Barretina J, Caponigro G, Stransky N, *et al*. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; **483**(7391): 603–7.
20. O'Neil J, Benita Y, Feldman I, *et al*. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016; **15**(6): 1155–62.
21. 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**(1): 1–17.
22. Liu H, Zhang W, Zou B, *et al*. Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020; **48**(D1): D871–81.
23. Sałat R, Sałat K. The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy. *Comput Methods Programs Biomed* 2013; **111**(2): 330–7.
24. Qi Y. Random forest for bioinformatics. In: *Ensemble machine learning*. Springer, 2012, 307–23.
25. Liu H, Zhang W, Nie L, *et al*. Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network. *BMC bioinformatics* 2019; **20**(1): 1–12.
26. Preuer K, Lewis RPI, Hochreiter S, *et al*. Deepsynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018; **34**(9): 1538–46.
27. Liu Q, Xie L. Transynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and path-

way deconvolution of drug combinations. *PLoS Comput Biol* 2021; **17**(2):e1008653.
28. Kuru HB, Tastan O, Cicek E. Matchmaker: A deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021.
29. Cao D-S, Xu Q-S, Hu Q-N, *et al*. Chemopy: freely available python package for computational biology and chemoinformatics. *Bioinformatics* 2013; **29**(8): 1092–4.
30. Yang W, Soares J, Greninger P, *et al*. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2012; **41**(D1): D955–61.
31. Wang T, Szedmak S, Wang H, *et al*. Modeling drug combination effects via latent tensor reconstruction bioRxiv. 2021.
32. Deng L, Cai Y, Zhang W, *et al*. Pathway-guided deep neural network toward interpretable and predictive modeling of drug sensitivity. *J Chem Inf Model* 2020; **60**(10): 4497–505.
33. Liu S, Tang B, Chen Q, *et al*. Drug-drug interaction extraction via convolutional neural networks. *Comput Math Methods Med* 2016; **2016**.
34. Wu Z, Ramsundar B, Feinberg EN, *et al*. Moleculenet: a benchmark for molecular machine learning. *Chem Sci* 2018; **9**(2): 513–30.
35. Xiong Z, Wang D, Liu X, *et al*. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem* 2019; **63**(16): 8749–60.
36. Sun Z, Huang S, Jiang P, *et al*. Dtf: Deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics* 2020; **36**(16): 4483–9.
37. Zhang T, Zhang L, Payne PRO, *et al*. Synergistic drug combination prediction by integrating multiomics data in deep learning models. In: *Translational Bioinformatics for Therapeutic Development*. Springer, 2021, 223–38.
38. Weininger D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988; **28**(1): 31–6.
39. 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**(D1): D1074–82.
40. Landrum G, *et al*. *Rdkit: Open-source cheminformatics*, 2006.
41. Loewe S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 1953; **3**:285–90.
42. Di Veroli GY, Fornari C, Wang D, *et al*. Combenefit: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics* 2016; **32**(18): 2866–8.
43. Ramsundar B, Eastman P, Walters P, *et al*. *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more*. O'Reilly Media, Inc, 2019.
44. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks arXiv preprint arXiv:1609.02907. 2016.
45. Cheng L, Li L. Systematic quality control analysis of lincs data. *CPT Pharmacometrics Syst Pharmacol* 2016; **5**(11): 588–98.
46. Derrien T, Johnson R, Bussotti G, *et al*. The gencode v7 catalog of human long noncoding rnas: analysis of their gene structure, evolution, and expression. *Genome Res* 2012; **22**(9): 1775–89.
47. Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemporary oncology* 2015; **19**(1A): A68.
48. Modjtahedi H, Cho BC, Michel MC, *et al*. A comprehensive review of the preclinical efficacy profile of the erbb family blocker afatinib in cancer. *Naunyn Schmiedebergs Arch Pharmacol* 2014; **387**(6): 505–21.
49. Silva-Oliveira RJ, Melendez M, Martinho O, *et al*. Akt can modulate the in vitro response of hnscc cells to irreversible egfr inhibitors. *Oncotarget* 2017; **8**(32):53288.
50. Hung M-S, Chen I-C, Lung J-H, *et al*. Epidermal growth factor receptor mutation enhances expression of cadherin-5 in lung cancer cells. *PLoS One* 2016; **11**(6):e0158395.
51. Bedard PL, Hyman DM, Davids MS, *et al*. Small molecules, big impact: 20 years of targeted therapy in oncology. *The Lancet* 2020; **395**(10229): 1078–88.
52. Goel S, Wang Q, Watt AC, *et al*. Overcoming therapeutic resistance in her2-positive breast cancers with cdk4/6 inhibitors. *Cancer Cell* 2016; **29**(3): 255–69.
53. Ye L, Mayerle J, Ziesch A, *et al*. The pi3k inhibitor copanlisib synergizes with sorafenib to induce cell death in hepatocellular carcinoma. *Cell death discovery* 2019; **5**(1): 1–12.
54. D'Alessandro R, Refolo MG, Lippolis C, *et al*. Modulation of regorafenib effects on hcc cell lines by epidermal growth factor. *Cancer Chemother Pharmacol* 2015; **75**(6): 1237–45.
55. Tang F, Li S, Liu D, *et al*. Sorafenib sensitizes melanoma cells to vemurafenib through ferroptosis. *Transl Cancer Res* 2020; **9**(3): 1584.
56. Zhang W-J, Li Y, Wei M-N, *et al*. Synergistic antitumor activity of regorafenib and lapatinib in preclinical models of human colorectal cancer. *Cancer Lett* 2017; **386**:100–9.
57. Lin X, Quan Z, Wang Z-J, *et al*. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In: *IJCAI*, Vol. **380**, 2020, 2739–45.
58. Zheng S, Rao J, Song Y, *et al*. Pharmkg: a dedicated knowledge graph benchmark for bomedical data mining. *Brief Bioinform* 2021; **22**(4):bbaa344.
59. Thafar MA, Olayan RS, Ashoor H, *et al*. Dtigems+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *J Chem* 2020; **12**(1): 1–17.
60. Zagidullin B, Wang Z, Guan Y, *et al*. Comparative analysis of molecular representations in prediction of drug combination effects bioRxiv. 2021.
61. Akhtar MJ. Covid19 inhibitors: a prospective therapeutics. *Bioorg Chem* 2020; **101**:104027.
62. Pereira TC, deMenezes RT, deOliveira HC, *et al*. In vitro synergistic effects of fluoxetine and paroxetine in combination with amphotericin b against cryptococcus neoformans. *Pathogens and Disease* 2021.
63. Ontong JC, Ozioma NF, Voravuthikunchai SP, *et al*. Synergistic antibacterial effects of colistin in combination with aminoglycoside, carbapenems, cephalosporins, fluoroquinolones, tetracyclines, fosfomycin, and piperacillin on multidrug resistant klebsiella pneumoniae isolates. *Plos one* 2021; **16**(1):e0244673.