

Data and text mining

Dual-view jointly learning improves personalized drug synergy prediction

Xueliang Li  ¹, Bihan Shen  ¹, Fangyoumin Feng  ¹, Kunshi Li¹, Zhixuan Tang¹, Liangxiao Ma  ², Hong Li  ^{1,*}

¹CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

²Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Science, Shanghai 200031, China

*Corresponding author: CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Yueyang Road 320, Shanghai. E-mail: lihong01@sinh.ac.cn

Associate Editor: Zhiyong Lu

Abstract

Motivation: Accurate and robust estimation of the synergistic drug combination is important for medicine precision. Although some computational methods have been developed, some predictions are still unreliable especially for the cross-dataset predictions, due to the complex mechanism of drug combinations and heterogeneity of cancer samples.

Results: We have proposed JointSyn that utilizes dual-view jointly learning to predict sample-specific effects of drug combination from drug and cell features. JointSyn outperforms existing state-of-the-art methods in predictive accuracy and robustness across various benchmarks. Each view of JointSyn captures drug synergy-related characteristics and makes complementary contributes to the final prediction of the drug combination. Moreover, JointSyn with fine-tuning improves its generalization ability to predict a novel drug combination or cancer sample using a small number of experimental measurements. We also used JointSyn to generate an estimated atlas of drug synergy for pan-cancer and explored the differential pattern among cancers. These results demonstrate the potential of JointSyn to predict drug synergy, supporting the development of personalized combinatorial therapies.

Availability and implementation: Source code and data are available at <https://github.com/LiHongCSBLab/JointSyn>.

1 Introduction

Traditional and modern medicines have always utilized the combinatorial drug therapies to better treat cancers (Fouquer and Guedj 2015). Compared with monotherapies, the drug combinations may improve treatment efficacy (Fisusi and Akala 2019, Sicklick *et al.* 2019), reduce toxicity and side effects (Al-Lazikani *et al.* 2012, Iqbal *et al.* 2022) and decrease the drug resistance (Tyers and Wright 2019, Liu *et al.* 2020). However, responses to the same drug combination may vary largely among patients due to the high heterogeneity of cancers (Sicklick *et al.* 2019). How to select suitable drug combination for an individual is a key challenge in personalized cancer therapy (Fan *et al.* 2021). With the development of high-throughput drug combination screening in recent years (He *et al.* 2018), the responses of drug combinations can be tested on multiple cancer cell lines simultaneously (Astashkina *et al.* 2012). However, exhaustive searching of the entire combination space is impossible due to high costs and time consumption for the exponential increase of potential combinations (Macarron *et al.* 2011, Goswami *et al.* 2015). Therefore, computational methods are needed to discover candidate synergistic drug combinations for experimental validation.

O’Neil *et al.* (2016) published a large-scale study, including 22 737 experimental measurements of 38 drugs on 39 cancer

cell lines. NCI published a larger study of the drug combination, including 304 549 measurements of 104 drugs on 60 cancer cell lines (Holbeck *et al.* 2017). Based on observed responses, a synergy score is calculated for each “drug1-drug2-cell line” triple and the triple can be classified as synergistic, antagonistic, or additive. DrugComb collected multiple experimental datasets and used the unified process to calculate synergy scores (Zagidullin *et al.* 2019, Zheng *et al.* 2021). In addition, pharmacogenomic databases such as Cancer Cell Line Encyclopedia (CCLE) provided comprehensive molecular measurement of cancer cell lines, including genomic mutations and copy number variations, RNA, and microRNA expression profiles (Ghandi *et al.* 2019). These resources provide a solid data basis for developing computational models to predict sample-specific drug synergy from molecular data.

Before 2018, some machine learning methods were used to predict drug synergy, such as Bayesian Network (Li *et al.* 2015), Logistic Regression (Li *et al.* 2020), Random Forest (Wildenhain *et al.* 2015, Li *et al.* 2017), and XGBoost (Janizek *et al.* 2018, Celebi *et al.* 2019, Sidorov *et al.* 2019). In recent years, more and more deep learning methods have been developed (Fan *et al.* 2021, Torkamannia *et al.* 2022). Preuer *et al.* (2018) built DeepSynergy, which used chemical information from the drugs and genomic information from

cell lines as input, and conical layers to model drug synergies. Sun *et al.* (2020) constructed DTF based on the tensor factorization method, which extracted latent features from drug synergy information. Jiang *et al.* (2020) proposed a graph convolutional network (GCN), which constructed cell line-specific drug-drug combination, drug-protein interaction, and protein-protein interaction networks. Zhang *et al.* (2021) built AuDNNsynergy which was trained using all tumor samples from The Cancer Genome Atlas. Kuru *et al.* (2022) proposed MatchMaker, which used drug chemical structure information and gene expression profiles of cell lines in a deep learning framework to predict drug synergy scores. Wang *et al.* (2022) proposed DeepDDS based on GCN, which transformed the drug to a molecular graph. Liu *et al.* (2022) proposed HypergraphSynergy, which converted the drug synergy task to the link prediction task and learned drug and cell line embeddings from hypergraphs. Hu *et al.* (2022) proposed DTSyn based on the transformer model, which extracted cell line expression profiles information with gene embedding (Hu *et al.* 2022). Pang *et al.* (2024) proposed a multimodal DL framework, MMSyn, which took the structure, fingerprint, and string encoding of drug molecules, gene expression, DNA copy number and pathway activity of cancer cell lines as input, and used attention mechanisms to integrate these data.

Although the above methods have achieved satisfactory performance with randomly splitting, but their accuracy decreased significantly for unseen drugs and cancer samples, especially for cross-study prediction (Wang *et al.* 2022). In a recent study, Zhang *et al.* found that the major cause of the variability between different studies was the experimental settings of drug dosage. They proposed a method to overcome experimental variability by harmonizing the dose-response curves of different studies (Zhang *et al.* 2023). The Personalized Deep Synergy Predictor (PDSP) proposed by Kuru *et al.* (2024) migrated the prediction of drug synergy scores from cell lines to the patient level. With the development of various transfer learning technologies in computational fields (Dalkiran *et al.* 2023, Huang *et al.* 2023), we supposed that fine-tuning might improve the drug synergy prediction on new drugs and cancer samples even for a small dataset. Additionally, current methods modeled two drugs independently, did not sufficiently utilize the association between two drugs. Multidrug representation learning has been shown to be efficient in predicting drug-drug interactions (Ren *et al.* 2022, Li *et al.* 2023). Through multi-drug joint representation, the information of the two drugs can communicate with each other, and the model can better learn the features of the two drugs interactively. The similar idea may also promote drug synergy prediction.

Based on these, we proposed a novel deep learning model JointSyn to predict the personalized synergistic effect of drug combination. It improves from previous models by dual-view joint representation of drugs and cell lines, also fine-tuning for better robustness. JointSyn is the overall best performer on the benchmark datasets for both regression and classification tasks, for new drugs or cell lines within a dataset or cross muti-datasets. Finally, the application of JointSyn to large-scale tumor cell lines obtains an estimated atlas of synergistic drug combinations for pan-cancer. Overall, both performance and case studies have proven that JointSyn is an effective tool for predicting the drug synergy, and our study also can provide quantitative suggestions for better experimental design.

2 Materials and methods

2.1 Data

2.1.1 Drug synergy dataset

We downloaded two large-scale drug synergy datasets (O'Neil, NCI-ALMANAC) from the DrugComb database. For each triplet (drug1-drug2-cell line), the synergy score was defined by loewe additivity (LOEWE) (Zagidullin *et al.* 2019). If a triple has inconsistent signs or higher coefficients of variation (>0.5) in synergy scores across replicates, we consider the measurements for these triples unreliable, so they are excluded. For the remaining triplets, we took the median of multiple measurements as the final synergy score of the triple. The final O'Neil dataset consists of 38 drugs, 34 cell lines, and 12 033 triplets (O'Neil *et al.* 2016); The NCI-ALMANAC dataset consists of 103 drugs, 46 cell lines, and 236190 triplets (Holbeck *et al.* 2017) (Supplementary Table S1). In order to make the model comparison more credible, we also used the drug combination sensitivity score (CSS value) as the response variable to observe the performance of the models (Malyutina *et al.* 2019).

2.1.2 Drug features

In order to better represent the molecular structure and physicochemical properties of the drug, we used the molecular graph and the Morgan fingerprint as drug features. SMILES of drugs were obtained from PubChem. First, RDKit was used to convert the SMILES into a molecular graph, in which the nodes are atoms and the edges are chemical bonds (Landrum). A 78D feature (Wang *et al.* 2022) was calculated for each node by DeepChem (Ramsundar *et al.* 2019). Secondly, RDKit was used to convert the SMILES into Morgan fingerprints with a radius of 6 (Preuer *et al.* 2018).

2.1.3 Cell line features

Gene expression profiles and somatic mutations of cancer cell lines were collected from the CCLE database. Transcripts per million (TPM) values were log2 transformed and normalized (Barretina *et al.* 2012). A previous work PaccMann reported 2128 drug sensitivity-related genes screened from expression profiles and PPI networks (Manica *et al.* 2019). Considering the intersection of 2128 genes and the CCLE gene expression profiles, expression values of 2087 genes were used as input features of cell lines.

2.2 JointSyn

2.2.1 Model architecture

We proposed a novel deep learning method named JointSyn to predict drug synergy (Fig. 1A). The input of JointSyn is the joint graph of the drug combination, the Morgan fingerprint of the two drugs, and the expression profile of the cell lines. JointSyn consists of two views: view 1 extracts the embedding of drug combination on cell lines, and view 2 contacts the combination of drug embedding on cell lines. Subsequently, the Prediction Net uses the embeddings from two views to predict the drug synergy of the drug combination on cell lines. More details about JointSyn are introduced below.

2.2.2 View 1: learning the embedding of drug combination on cell lines

The molecular graph of a drug is defined as $G = (V, E)$, where V is a set of N nodes represented by vectors, and the i th atom can be represented as $v_i \in V$; E is a set of edges. The chemical bond between the i th and j th atoms can be

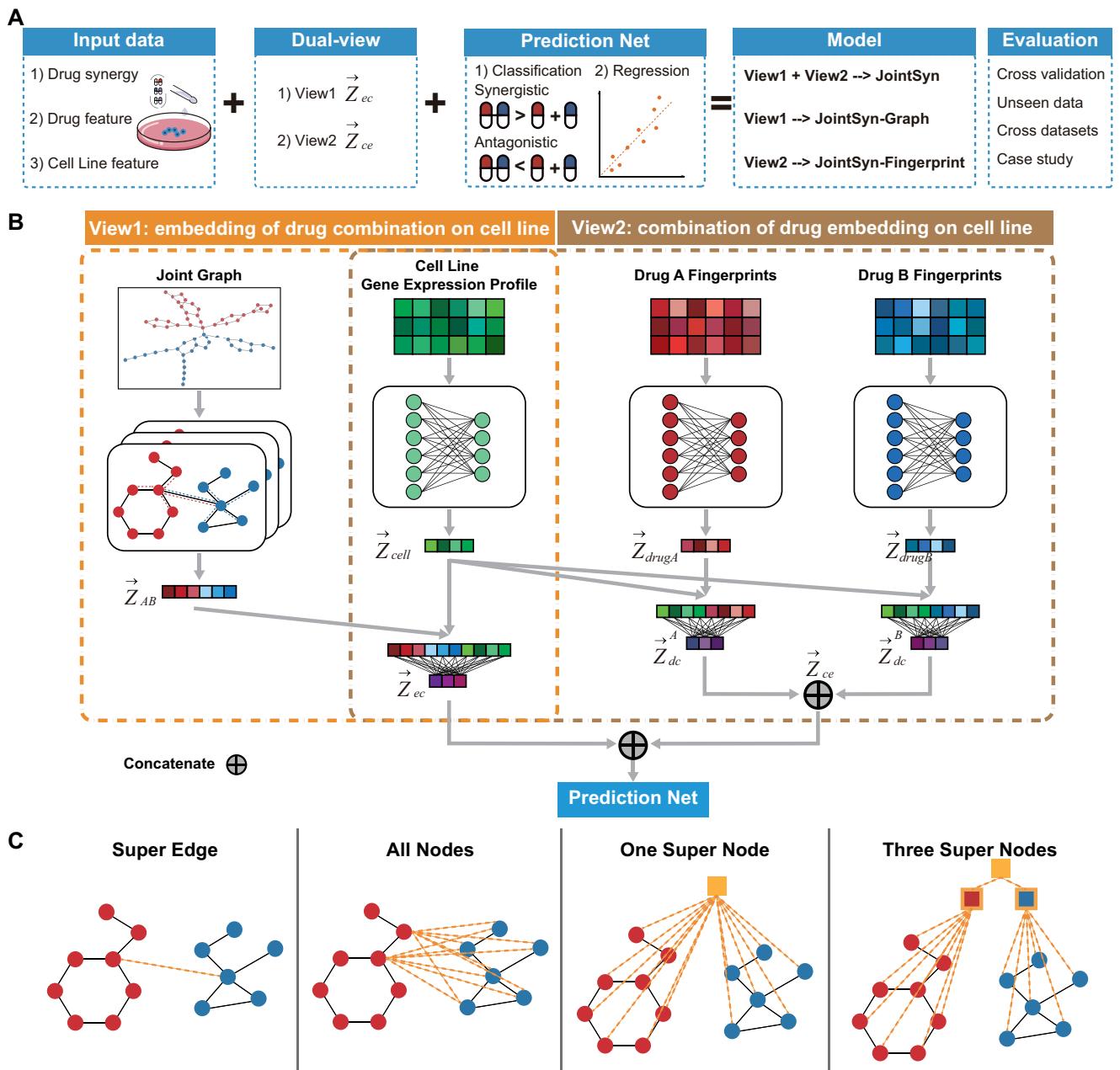


Figure 1. Overview of JointSyn. (A) The schematic diagram of JointSyn. JointSyn uses drug synergy, drug features and cell line features as input, and extracts drug synergy-related features from two views on classification and regression tasks. JointSyn is evaluated by cross-validation, unseen data, cross-datasets and case study. (B) JointSyn consists of dual views to capture the drug synergy-related features. The drug combination embedding processed through GAT based on the joint graph concatenates with the gene expression embedding to learn the embedding of drug combination on cell lines. Each drug's embedding is processed through MLP based on the morgan fingerprint and concatenated with the gene expression embedding, then two drugs' results are concatenated to learn the combination of drug embedding on cell lines. (C) Four methods for constructing the joint graph of drug combination.

expressed as $e_{i,j} \in E$, or it can be expressed as $\langle v_i, v_j \rangle \in E$. In this view, we want to get the embedding of drug combination on cell lines, so the first task is to get a joint graph representation of the drug combination. As shown in Fig. 1B, four methods were proposed to get the joint graph $G_{AB} = (V_{AB}, E_{AB})$ from drug A G_A and drug B G_B .

Super-Edge: We firstly calculated betweenness centrality to measure the importance of a node in connecting other nodes in each drug molecular graph:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}. \quad (1)$$

Among them, σ_{st} is the number of shortest paths from node s to t , and $\sigma_{st}(v)$ is the number of paths through node v among these shortest paths.

Then we selected the two nodes with the highest betweenness centrality from two drug graphs, namely $v_1 = \arg \max_{v \in V_A} C_B(v)$ and $v_2 = \arg \max_{v \in V_B} C_B(v)$, virtually added an edge between

two nodes. Currently, $V_{AB} = V_A \cup V_B$, $E_{AB} = E_A \cup E_B \cup \{(v_1, v_2)\}$. The virtual edge can transfer atomic information between drugs and the model can capture cross-drug interactions.

All-nodes: We constructed a bipartite graph G_{AB} from the molecular graphs of G_A and G_B , where $G_{AB} = G_A \times G_B$, $G_{AB} = (\overline{V}_{AB}, \overline{E}_{AB})$. Specifically, each atom in the drug G_A builds edges connection with each atom in drug G_B . Currently, $V_{AB} = V_A \cup V_B$, $E_{AB} = E_A \cup E_B \cup \overline{E}_{AB}$.

One-super-node: We defined a super node which was connected to each atom in the two drugs:

$$V_{AB} = V_A \cup V_B \cup \{v_{Sup}\}, \quad (2)$$

$$E_{AB} = E_A \cup E_B \bigcup_{i=1}^{n_A} (v_i, v_{Sup}) \bigcup_{j=1}^{n_B} (v_j, v_{Sup}), \quad (3)$$

where n_A and n_B are the numbers of atoms in drug A and B, respectively.

Three-super-nodes: Super node 1 (*Sup1*) is connected to all atoms in drug A, and super node 2 (*Sup2*) is connected to all atoms in drug B. *Sup1* and *Sup2* are aggregated by *Sup*.

$$\begin{aligned} V_{AB} &= V_A \cup V_B \cup \{v_{Sup1}, v_{Sup2}, v_{Sup}\}. \\ E_{AB} &= E_A \cup E_B \cup \{(v_{Sup1}, v_{Sup}), (v_{Sup2}, v_{Sup})\} \\ &\quad \bigcup_{i=1}^{n_A} (v_i, v_{Sup1}) \bigcup_{j=1}^{n_B} (v_j, v_{Sup2}). \end{aligned} \quad (4) \quad (5)$$

Extracting drug combination features based on graph attention network: We used the graph attention network (GAT) to extract the embedding of drug combination from the joint graph $G_{AB} = (V_{AB}, E_{AB})$, whose node feature matrix is X and the adjacency matrix is A . The output features of the nodes after each layer of iterative propagation are as follows:

$$\vec{h}'_i = \parallel_{k=1,2,\dots,n} \sigma \left(\sum_{j \in N_i} \alpha_{ij}^{(k)} W^{(k)} \vec{h}_j \right), \quad (6)$$

where k is the number of attention heads, \parallel concatenates the outputs of multiple attention mechanisms, and $W^{(k)}$ is the weight matrix that can be learned. The attention coefficient $\alpha_{ij}^{(k)}$ between each atom i and its neighbor atom j is calculated as follows:

$$\alpha_{ij}^{(k)} = \frac{\exp \left(\text{LeakyReLU} \left(\vec{a}^T [W^{(k)} \vec{h}_i \parallel W^{(k)} \vec{h}_j] \right) \right)}{\sum_{l \in N_i} \exp \left(\text{LeakyReLU} \left(\vec{a}^T [W^{(k)} \vec{h}_i \parallel W^{(k)} \vec{h}_l] \right) \right)}, \quad (7)$$

where $W^{(k)}$ is the weight matrix shared with the above, the purpose is to enhance the features of vertices, a \vec{a}^T is the weight vector that can be learned.

Our model is based on a three-layer GAT: each atom can see its three-hop neighbors, and atomic information can be transferred between drugs. After the last layer of GAT, we add a global pooling layer to aggregate the learned atomic features to obtain the embedding \vec{z}_{AB} of the drug combination.

Extracting cell line features based on multi-layer perception: We used a two-layer multi-layer perception (MLP) to construct embedding for the cell line as follows:

$$\vec{z}_{cell}^l = \text{ReLU} \left(W_{cell}^l \vec{z}_{cell}^{l-1} + b_{cell}^l \right), \quad (8)$$

where W_{cell} and b_{cell} are weight matrices that can be learned, l means in the l -layer. $\vec{z}_{cell}^0 = X_{cell}$ is the expression profile of 2087 genes. Finally, the embedding \vec{z}_{cell} of the cell line is obtained.

Concatenating the embedding of drug combinations on cell lines: The embedding of drug combination on cell line \vec{z}_{ec} can be obtained through the fully connected layer:

$$\vec{z}_{ec} = \text{ReLU} \left(W_{ec} [\vec{z}_{AB} \parallel \vec{z}_{cell}] + b_{ec} \right), \quad (9)$$

where W_{ec} and b_{ec} are weight matrices that can be learned.

2.2.3 View 2: learning the combination of drug embeddings on cell lines

Each drug's embedding \vec{z}_{drug} from Morgan fingerprint feature was concatenated with the cell line embedding \vec{z}_{cell} , and then inputted into MLP to obtain the embedding of one drug on cell line \vec{z}_{dc} :

$$\vec{z}_{dc} = \text{ReLU} \left(W_{dc} [\vec{z}_{cell} \parallel \vec{z}_{drug}] + b_{dc} \right). \quad (10)$$

Then we can get the combination of drug embedding on cell line \vec{z}_{ce} :

$$\vec{z}_{ce} = \text{ReLU} \left(W_{ce} [\vec{z}_{dc}^A \parallel \vec{z}_{dc}^B] + b_{ce} \right). \quad (11)$$

2.2.4 Prediction net: predicting drug synergy based on dual-view

Through the above two networks, we can get the dual-view embedding about drug synergy. By contacting these two embeddings and inputting them into the three-layer MLP:

$$\vec{z}^l = \text{ReLU} \left(W^l \vec{z}^{l-1} + b^l \right), \quad (12)$$

where l means in the l -layer. At the first layer, $\vec{z}^0 = [\vec{z}_{ec} \parallel \vec{z}_{ce}]$, and the final embedding $\vec{z} = \vec{z}^3$.

Through the embedding \vec{z} of drug synergy, we can get the final predicted value:

$$out = \sigma(W_{out} \vec{z} + b_{out}), \quad (13)$$

where σ is the softmax or linear activation function for the classification or regression task.

2.2.5 Global parameters

The architecture of JointSyn is determined by many hyperparameters, including but not limited to loss function, activation function, learning rate and regularization method. We used grid search to adjust the hyperparameters and performed ten times of 5-fold cross-validation to increase the robustness. Detailed parameter settings are in [Supplementary Table S2](#).

2.3 Method evaluation

We compared JointSyn with several state-of-the-art methods (DeepSynergy, AuDNNsynergy, DeepDDS, DTsyn, Hypergraph Synergy, Matchmaker, MMSyn, and XGBoost) by 5-fold cross

validation. Parameter selections of different compared methods were undergone according to the instructions of the original papers. In the regression task, we adopted Pearson Correlation Coefficient (PCC) and R square (R2) as the main evaluation metric, and we also reported other widely used performance metrics, including Mean Squared Error (MSE) and Root Mean Square Error (RMSE). In the classification task, we adopted Kappa and F1 as the main metrics, and we also reported other classic performance metrics, including Area under the ROC Curve (ROC AUC), Precision-Recall Curve (PR AUC), Balanced Accuracy (BACC), Precision, and Recall. We performed ten times 5-fold cross-validations randomly and reported the mean and 95% confidence intervals of the performance metrics.

To ensure the generalization ability of the model on unseen data, four data splitting strategies were used and compared: Random splitting that divided all triplets into five folds; PairOut splitting that randomly divided all drug combinations into five folds to ensure that the drug combinations in test set do not appear in the training set; CellOut splitting that randomly divided all cell lines into five folds to ensure that the cell lines in test set do not appear in the training set; DrugOut splitting that randomly divided all drugs into five folds to ensure that the drugs in test set do not appear in the training set.

3 Results

The goal of JointSyn is to more accurately and robustly estimate the personalized synergistic effects of drug combinations. Its schematic diagram is shown in Fig. 1A. Its core features are to get the global graph information on drug combinations by joint representation, and to construct two views to capture the drug synergy-related features as shown in Fig. 1B. One view is the embedding of drug combination on cancer cell lines (view1: drugCombination-cell), and the other is the combination of two drugs' embeddings on cancer cell lines (view2: drug1-cell & drug2-cell). In view1, molecular graphs of two drugs are firstly joined to construct a graph of the drug combination; next GAT is used to obtain a joint representation of the drug combinations; then the representation of the drug combination is concatenated the gene expression profiles of cell lines. The rationale of view1 is to better get the global graph information of drug combinations. In view2, each drug's molecular fingerprint and cell line's gene expression profile are integrated through MLP to get the embedding of each drug on cell lines respectively, and then the embeddings of two drugs were contacted to obtain the combination of drug embeddings on cell lines. Finally, JointSyn takes the embeddings from dual views to predict continuous or categorical synergy scores by MLP. The model only using view1 or view2 embedding and connecting MLP is named JointSyn-Graph or JointSyn-Fingerprint respectively.

We tested four joint methods for connecting drugs' molecular graphs: Super-Edge, All-Nodes, One-Super-Node, and Three-Super-Nodes (Fig. 1C, seeing methods for details). The Super-Edge strategy achieves the best performance on the O'Neil dataset (Supplementary Table S3). Therefore, we selected Super-Edge as the Joint Method in the following sections.

3.1 JointSyn improves drug synergy predictions

We first compared JointSyn with five state-of-the-art deep learning methods and one classic machine learning method

on two tasks (regression and classification) using the O'Neil dataset. JointSyn achieves the best performance on all evaluation indicators for the regression task (Fig. 2A, Supplementary Table S4). JointSyn's R2 is 0.78, which is 8.9% higher than the DeepSynergy which is the first method using deep learning for predicting drug synergy, and 6.1% higher than the Matchmaker which is the best baseline method. And JointSyn's PCC is 0.89, which is 5.3% higher than the DeepSynergy and 3.7% higher than the Matchmaker. When treating the drug synergy prediction as a classification task, JointSyn is still significantly better than other methods for Kappa and F1 scores, although the difference of ROC-AUC among JointSyn, XGBoost, DeepDDS and MMSyn is not obvious (Fig. 2B). The similar performance improvement of JointSyn is also observed on another benchmark dataset NCI-ALMANAC (Supplementary Fig. S1 and Supplementary Table S5). Overall, these results show the powerful ability of JointSyn to predict the drug synergy.

There are 583 drug combinations in the O'Neil dataset, and each drug combination was measured on multiple cell lines. We noted that there were large differences in the performance of JointSyn for different drug combinations, therefore we further explored the possible effect factors. When a drug combination is measured in more cell lines, JointSyn's prediction is more related with actual synergy scores (Fig. 2C). When the standard deviation of the synergy scores for a drug combination in multiple cell lines is larger, the correlation between real and predicted synergy scores is significantly higher (Fig. 2D). In other words, this means that if the synergy scores of a drug combination in multiple cell lines are significantly different, the synergy scores can be predicted more accurately.

3.2 Dual-view of JointSyn captures different aspects and achieve better performance

To inspect how well JointSyn generated the embeddings related with drug combinations on cell lines, we used t-distributed stochastic neighbor embedding (tSNE) to visualize the dual-view embeddings from the last layer of JointSyn (Fig. 3A). Similar plots were conducted using the single view of "drugCombination-cell" embedding in the JointSyn-Graph (Fig. 3B), and "drug1-cell & drug2-cell" embedding in the JointSyn-Fingerprint model (Fig. 3C). The synergistic and antagonistic triples can be well distinguished in the embedding space, and the dual-view seems better than the single view. These results indicate that dual-view representation of JointSyn is successful in extracting low-dimension embeddings related with drug synergy.

We further inspected the contribution of two views by comparing them on each drug combination. The number of drug combinations with PCC >0.7 for JointSyn is 424, for JointSyn-Fingerprint is 363, and for JointSyn-Graph is 332 (Fig. 3D). These numbers are larger than baseline methods. Next, we selected the best method (with highest PCC) for each drug combination and counted the number of drug combinations that each method obtained the best performance. The JointSyn method achieves highest PCC for 314 drug combinations, while JointSyn-Fingerprint is 67 and JointSyn-Graph is 90 (Fig. 3E). Supplementary Figure S2 shows the top 20 drug combinations with the highest PCC by the JointSyn, JointSyn-Graph and JointSyn-Fingerprint respectively. The best performance on some drug combinations may be achieved when using only one view. More

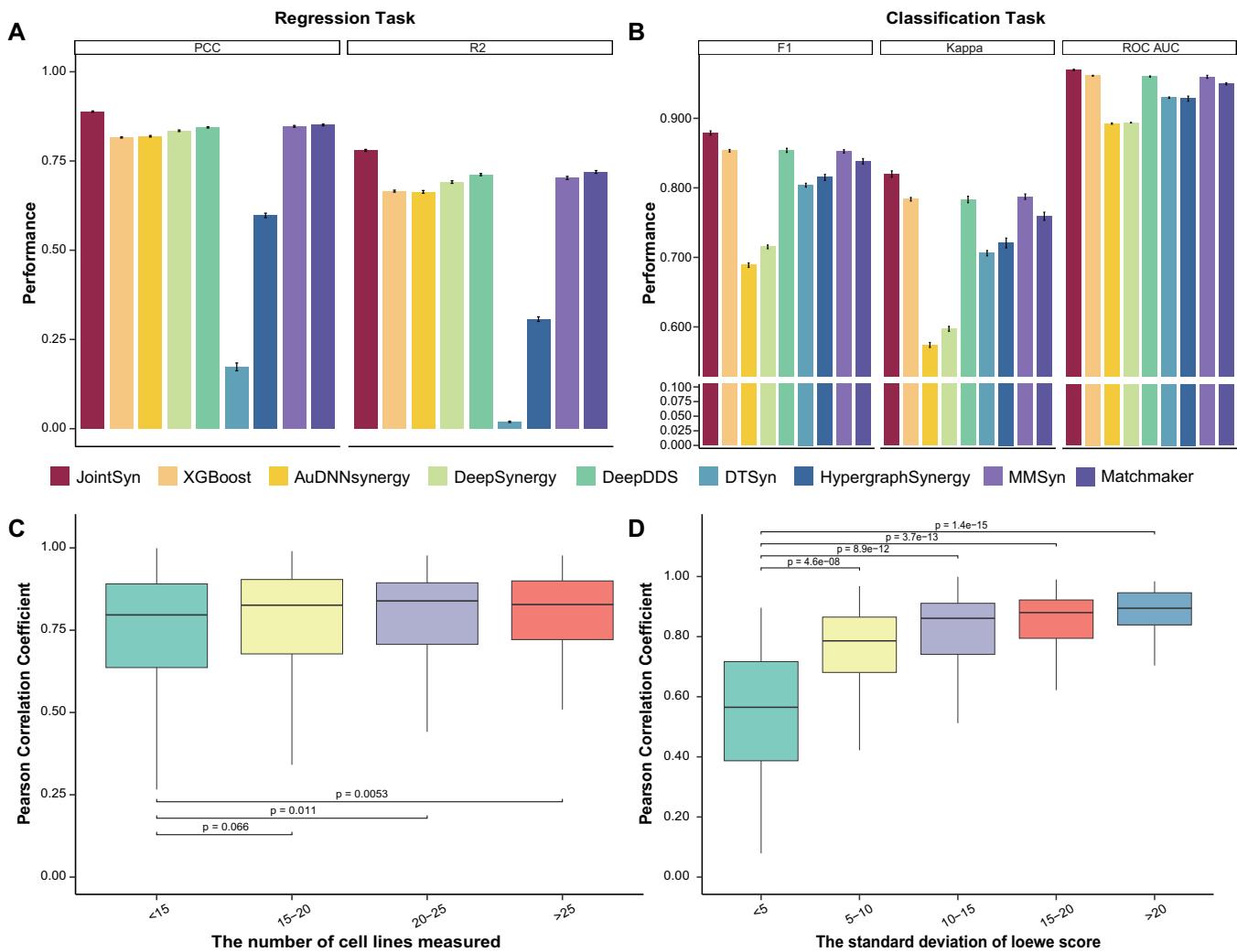


Figure 2. Evaluation of JointSyn using the O'Neil benchmark dataset. (A–B) Performance comparison of JointSyn and other methods for the regression and classification tasks. Five-fold cross-validations were replicated 10 times to calculate the standard deviations (error bars). (C–D) Taken PCC of O'Neil regression model as an example to discuss factors associated with the performance of each drug combination. (C) The relationship between the PCC and the number of cell lines measured for each drug combination. (D) The relationship between the PCC and the standard deviation of the real synergy scores.

importantly, dual-view achieves the best performance for most drug combinations due to the capture of complementary aspects of embedding spaces.

3.3 JointSyn fine-tuning improves drug synergy predictions for unseen data

In the previous part, we evaluated the performance of JointSyn by randomized data splitting, but the real prediction tasks may involve unseen drug combinations, drugs, or cell lines. Therefore, we further evaluated JointSyn by three stratified data splitting scenarios in which the drug combinations, drugs, or cell lines used for prediction were not included in training dataset, named PairOut, DrugOut, and CellOut, respectively.

Figure 4A shows the performance of each method in four scenarios on the O'Neil dataset (Supplementary Table S6 for detailed results). Compared with randomized data splitting, the performance of all methods significantly decreased for stratified data splitting. DrugOut decreases the most, followed by CellOut and PairOut. For the PairOut scenario, models may still learn from other combinations that share a drug; for CellOut scenario, models may still learn from

similar cell lines; however, DrugOut simulates the prediction of a completely new drug, the lack of information in the training set resulting in a significant performance decline. Compared with other methods, JointSyn still achieves the best performance in most metrics in the stratified scenarios. For an unseen drug combination, JointSyn can averagely achieve a F1 of 0.84 and a PCC of 0.86; for an unseen cell line, JointSyn can still averagely achieve a F1 of 0.75 and a PCC of 0.67. The similar performance improvement of JointSyn was observed when CSS values were used as the synergy scores (Supplementary Table S7). JointSyn achieved excellent performance whether considering LOEWE only or simultaneously LOEWE and CSS, indicating JointSyn an effective tool for predicting drug combination efficacy (Supplementary Fig. S3).

Prediction of drug synergy for unseen data is very challenging. To address this challenge, we used the fine-tuning method to improve the JointSyn's performance by introducing a small number of experimental measurements (Fig. 4B). For a new cell line in CellOut splitting scenario, k drug combinations on this cell line were gradually added into the training set (k -shots). As the number of shots increased, PCC and R2 gradually increased and tended to be stable (Fig. 4C).

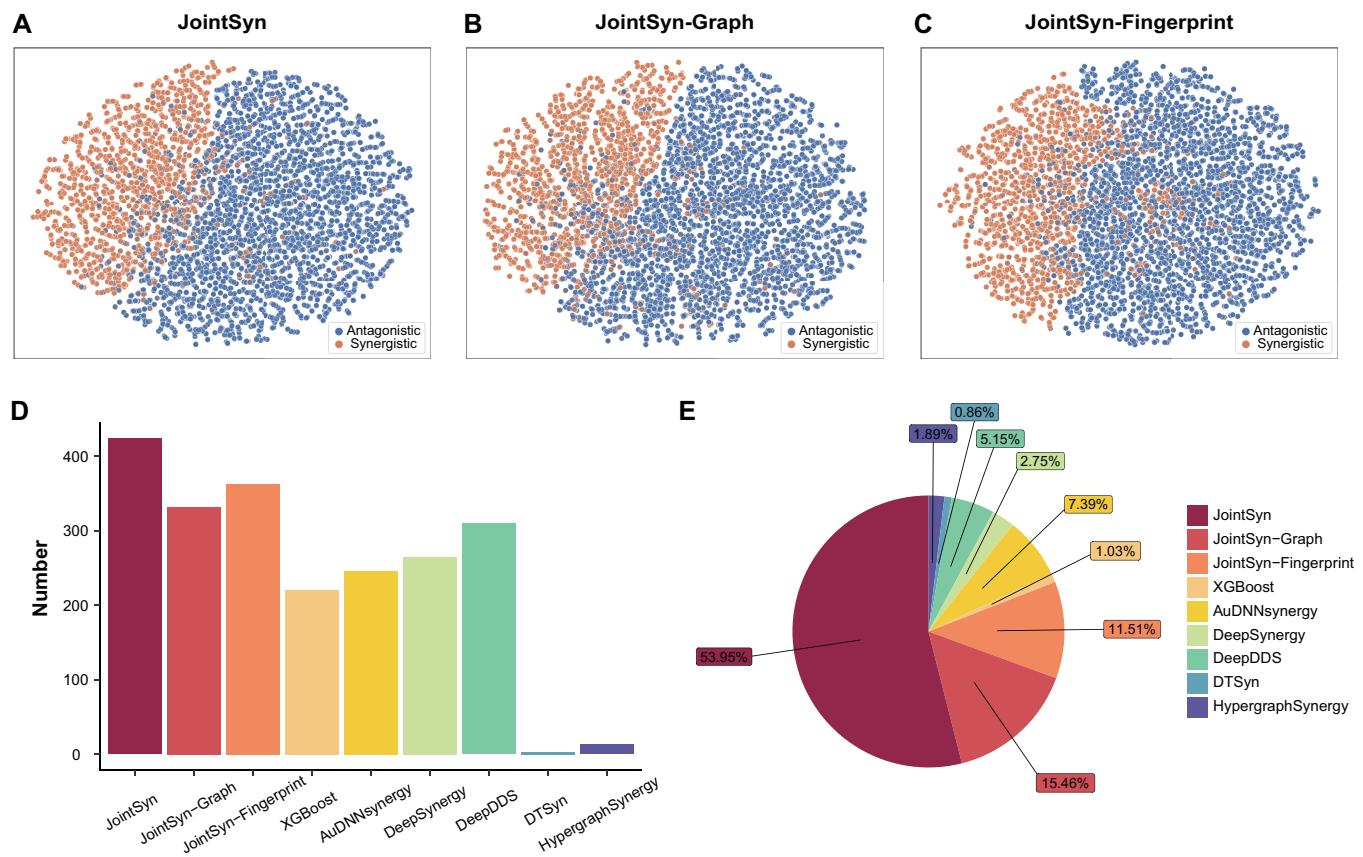


Figure 3. Evaluation the dual views of JointSyn. (A–C) TSNE plots show the last layer’s embedding from JointSyn (A), JointSyn-Graph (B), and JointSyn-Fingerprint (C). Each point is a triplet (drug1-drug2-cell line). Colors indicate synergistic and antagonistic triplets. (D–E) Performance metrics were calculated for each drug combination separately instead of mixing all samples together. (D) The number of drug combinations with $PCC > 0.7$ for each method. (E) The number of drug combinations for each method achieving the best performance.

The performance metrics at 150-shots is close to random splitting. This means for a given drug list, if 21% ($150/703 \approx 21\%$) combinations are experimentally measured for a cell line, and the synergy of the remaining drug combinations can be well predicted. Similarly, k -shots fine-tuning promoted JointSyn’s performance in the PairOut scenario (Fig. 4D). For a drug combination not in the training set, when the synergy scores of this combination on 15% ($5/34 \approx 15\%$) cell lines are added into the training set, JointSyn can well predict synergy on the remaining cell lines. Taken together, JointSyn with fine-tuning further improves its generalization ability for prediction.

3.4 JointSyn fine-tuning improves drug synergy predictions for independent datasets and provides quantitative suggestions for better experimental design

We further evaluated JointSyn for cross-dataset prediction using NCI-ALMANAC and O’Neil datasets. These 2 datasets only shared 14 drugs, 9 cell lines, and 221 triplets (drug1-drug2-cell line) (Fig. 5A). For the overlapped triplets, whose synergy scores were measured in both datasets, the correlation of two datasets is only 0.32 (Fig. 5B). Such low correlation may come from differences in experimental protocols between two laboratories such as the concentration range of drugs and the cell viability assays (Zhang *et al.* 2023). Taken the larger dataset NCI-ALMANAC (236190 triplets) as training data and the O’Neil dataset (12 033 triplets) as

independent testing data, the performance of JointSyn ($R^2 = 0.064$ and $PCC = 0.14$) and other baseline methods significantly dropped. Such low predictive performance indicates that direct application of the model to independent datasets is infeasible due to huge differences in experimental subjects and protocols.

The cross-dataset prediction is a common demand in real-world application scenarios and a very challenging task for computation modeling from limited training data, but previous methods did not pay enough attention to this issue. Therefore, we quantitatively explored how much data were needed for reliable predictions when training JointSyn by a prior public dataset and fine-tuning by a small independent dataset and compared with PDSP (Fig. 5C). Figure 5D shows the performance of the model fine-tuned by the O’Neil dataset based on the model trained by NCI-ALMANAC. For a given drug list, when 28% ($200/703 \approx 28\%$) drug combinations for each cell line of the O’Neil dataset are added to fine-tuned JointSyn, synergy scores of the remaining triplets can be well predicted ($R^2 = 0.75$, $PCC = 0.87$). Similarly, when 29% ($10/34 \approx 29\%$) cell lines for each drug combination are added, JointSyn can well predict synergy scores on the remaining cell lines ($R^2 = 0.73$, $PCC = 0.86$). Additionally, we compared JointSyn with another transferring learning method PDSP (Kuru *et al.* 2024). Although PDSP’s performance also improves after fine-tuning, JointSyn still outperforms PDSP when fine-tuned on the same data. This experiment proves that through fine-tuning with a small

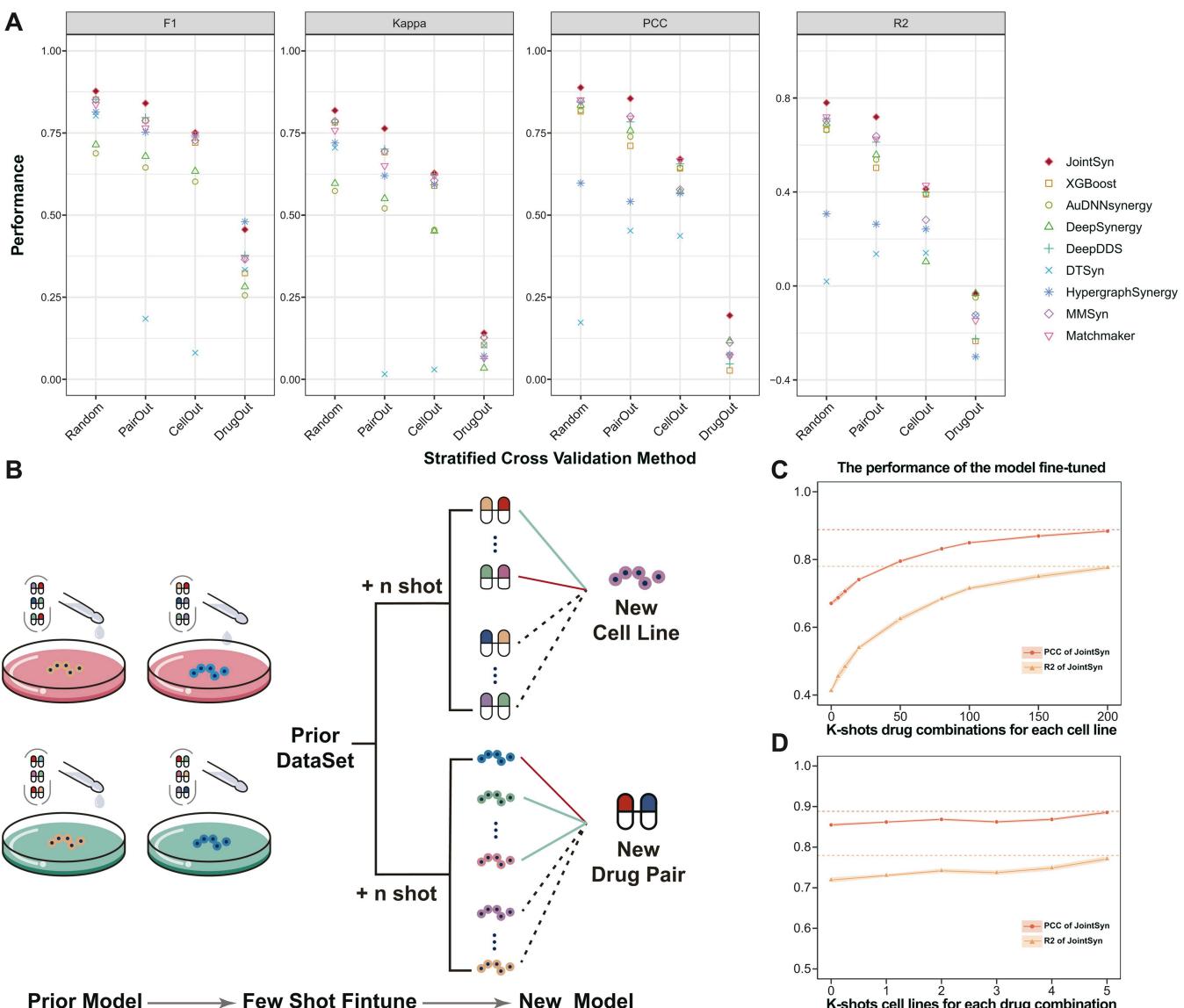


Figure 4. Model performance on unseen drug combinations, cell lines, and drugs. (A) Performance comparison for the regression and classification tasks in four data splitting scenarios of the O’Neil dataset. Random means data was randomly splitted into training and test set. Three stratified data splitting PairOut, CellOut, DrugOut means drug combinations, drugs, or cell lines that were used for prediction were not included in training dataset. (B) For a new cell line (CellOut) or drug combination (PairOut), JointSyn uses few-shot finetuning to improve performance. (C) The improvement of JointSyn performance after adding experimental measurements from different numbers of drug combinations when the cell line is new. The dashed line is the JointSyn’s performance in random data splitting scenarios. (D) The improvement of JointSyn performance after adding experimental measurements from different numbers of cell lines when the drug combination is new.

number of experimental measurements from the external dataset, the decrease of model performance on cross-dataset prediction can be solved.

3.5 Application of JointSyn to investigate synergistic drug combinations of pan-cancers

To further demonstrate the utility of JointSyn in personalized drug synergy prediction and explore the patterns of drug synergy among different tumors, we applied JointSyn trained with the O’Neil dataset to score 996 cell lines from seven tumor lineages of CCLE on 703 drug combinations. This generated an estimated atlas of synergistic drug combinations for pan-cancer (Fig. 6A). Ward’s hierarchical clustering divided tumors into 4 clusters (C1–C4) and drug combinations into 5 clusters (D1–D5) from the predicted drug synergy matrix. D1 and D2 have no synergistic effects in almost all tumors, while

other drug combination clusters show heterogeneous patterns of synergy scores among different tumors. Tumors in C1 are mainly composed of blood cancers and lymphomas, harboring no synergistic drug combinations, which may reflect huge differences between hematolymphoid tumors and solid tumors; C3 is mainly composed of skin cancer; Other types of solid tumors are mixed in C2 and C4.

Lung cancer has the largest number of cell lines in CCLE dataset. Next, we took lung cancer as an example to illustrate the heterogeneity among cancer samples of the same lineage. The drug synergy matrix for all 703 drug combinations in 188 lung cancer cell lines was illustrated in Supplementary Fig. S4A. We selected 108 drug combinations with a synergistic ratio >5% and drew a heatmap of synergy scores (Fig. 6B). Samples were divided into three clusters (Lung_C1~Lung_C3) based on the synergy scores of these

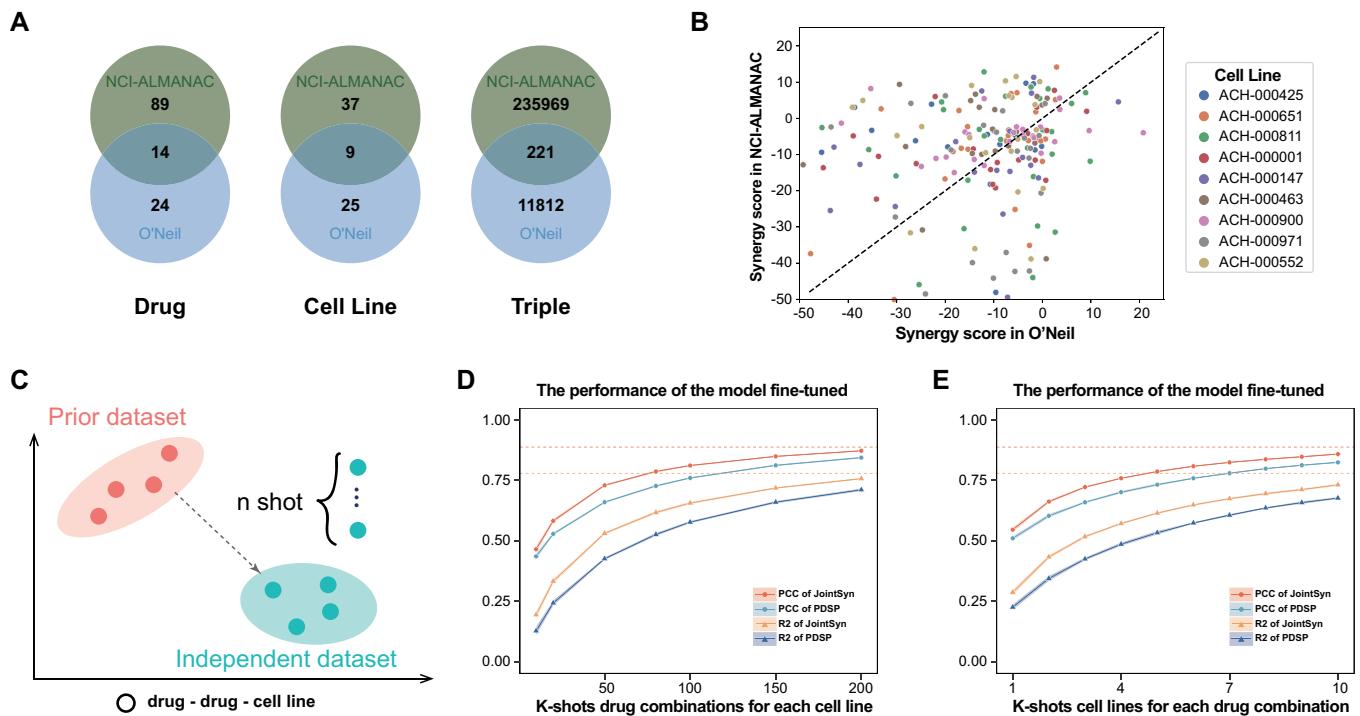


Figure 5. The difficulties and model performance for cross-dataset prediction. (A) Venn diagram of drugs, cell lines and triples (drug1-drug2-cell) in O'Neil and ALMANAC datasets. (B) A scatter plot shows the synergy scores of the overlapped measured triplets between O'Neil and NCI-ALMANAC datasets. (C) The process of few-shot finetuning for cross-dataset. (D–E) The performance of the model fine-tuned by the O'Neil dataset based on the model trained by NCI-ALMANAC dataset. (D) Performance improvement with the adding of experimentally measured drug combinations on a new cell line. (E) Performance improvement with the adding of experimentally measured cell lines on a new drug combination.

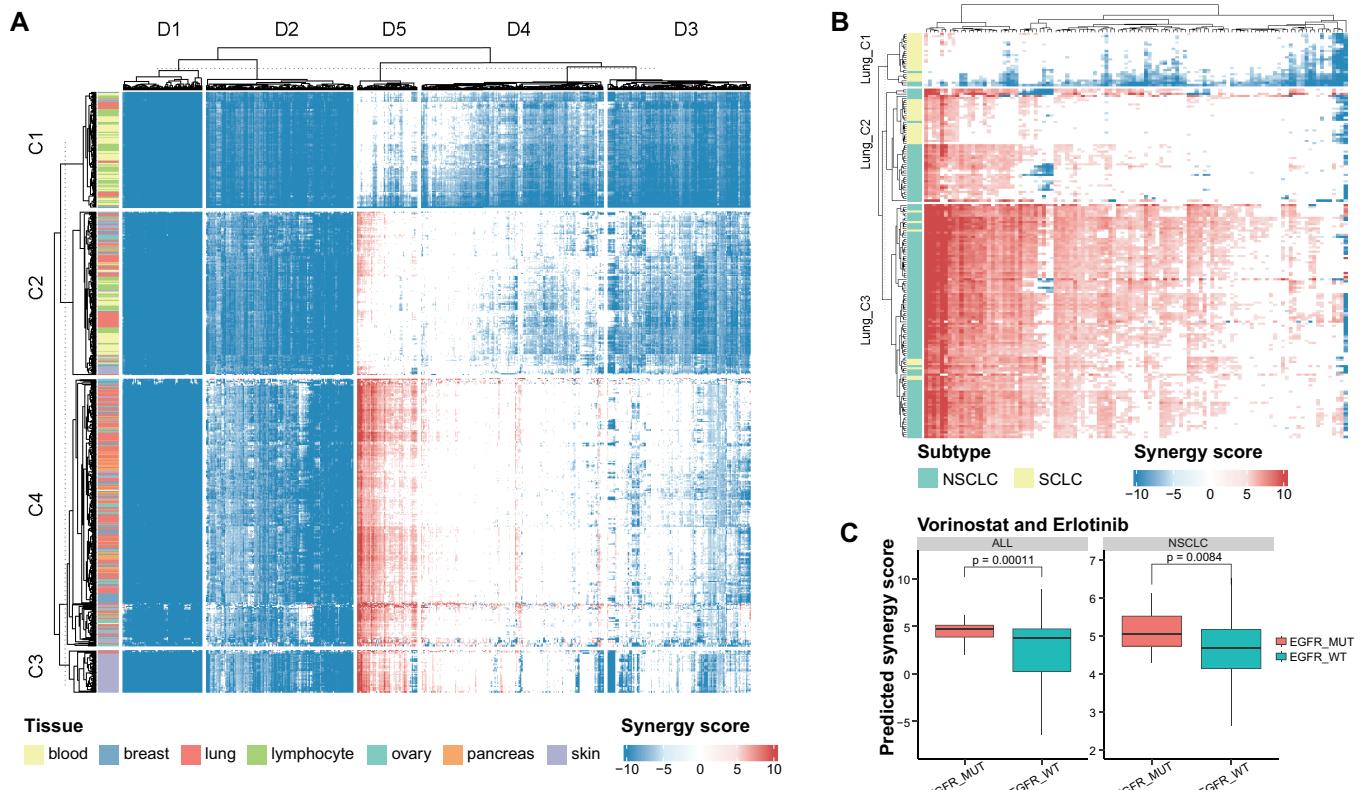


Figure 6. The predicted synergy scores of pan-cancers. (A) The predicted synergy scores of 703 drug combinations on 996 cell lines. Larger positive value indicates more synergism, and smaller negative value indicates more antagonism. (B) The predicted synergy scores of selected drug combinations on lung cancer cell lines. NSCLC: non-small cell lung cancer. SCLC: small cell lung cancer. (C) Comparison of the predicted synergy scores of vorinostat and erlotinib in cell lines with or without EGFR mutations. Pan-cancers and NSCLC were compared respectively. *P*-value was calculated by the Wilcoxon test.

selected drug combinations. The two main types of lung cancer, non-small cell lung cancer (NSCLC), and small cell lung cancer (SCLC), show obvious differences in the distribution of synergy scores. Lung_C1 is mainly composed of SCLC and has almost no synergistic drug combinations. Lung_C3 is mainly composed of NSCLC and many drug combinations have synergistic effects. This is consistent with many previous studies which mentioned SCLC is highly drug-resistant and has poor prognosis (Koinis *et al.* 2016, Schneider and Kalemkerian 2016). Some drug combinations have synergistic effects on specific NSCLC cell lines, therefore we further explored whether the differences of drug synergy were related to certain somatic mutations. EGFR is the frequently mutated driver gene in NSCLC and its inhibitor erlotinib is in our training dataset. Combinations of erlotinib with some other drugs are synergistic based on the predictions from JointSyn and associated with EGFR mutations. Taken “erlotinib and vorinostat,” “erlotinib and dactolisib,” “erlotinib and MK-2206” as examples, they have significantly high synergy scores in EGFR mutated cell lines than those without EGFR mutations, whether in NSCLC cell lines or in all cell lines (Fig. 6C, Supplementary Fig. S4B and C). Evidence related to these predicted drug synergies can be found in previous studies. Vorinostat can enhance the therapeutic potential of erlotinib in lung cancer cells (Alqosaibi *et al.* 2022); the combination of MK-2206 and erlotinib can synergistically inhibit the cell proliferation of human cancer cell lines (Hirai *et al.* 2010); PI3K/Akt/mTOR signaling is the main mechanism of EGFR resistance, and dactolisib is a dual PI3K/mTOR inhibitor (Wu *et al.* 2019), which may explain the synergy effect of erlotinib and dactolisib. In summary, the predictions from JointSyn are mostly reliable and some are supported by existing experimental evidence.

4 Discussion

In this work, we have proposed a novel deep learning model, JointSyn to predict drug synergy from dual-view jointly learning. JointSyn performs best compared to other methods on two benchmark datasets. The embedding from dual view has been proved to be significantly helpful in drug synergy prediction. More importantly, JointSyn utilizes few experimental measurements to fine-tune, improve its performance not only for the unseen subset within a dataset but also for independent datasets. Finally, an estimated atlas of synergistic drug combination for pan-cancer was generated by JointSyn and the differential patterns among tumors were discussed.

A common bottleneck of developing drug synergy prediction methods is the limited number of experimental measured synergy scores. A future direction is to incorporate large-scale unsupervised pre-training into the training process, so that the model can learn more drugs and cell lines even though these are unlabeled. JointSyn currently only uses the drug's molecular graph, Morgan fingerprint and the expression profile of cell line to predict drug synergy. More prior information such as drug-target genes, drug-drug interactions, and drug-disturbed expression profiles may also be useful. Incorporation more information may more sufficiently model drugs and cell lines, even better capture the association between drugs in the training set and novel drugs. In addition, drug dosage is important for improving the accuracy of cross-study transfer learning (Zhang *et al.* 2023, Tang *et al.* 2024), but existing studies rarely involve drug dosage

information. We consider using a new transfer learning method to improve JointSyn and incorporate drug dosage into the model to better solve the problem of cross-study drug synergy prediction. We will improve JointSyn from these perspectives in future. Overall, we believe that JointSyn is a valuable tool for pre-screening of synergistic drug combinations.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

The authors declare no competing financial interest.

Funding

This work was supported by National Natural Science Foundation of China [32170680, T2122018], Natural Science Foundation of Shanghai [21ZR1476000], and CAS Youth Innovation Promotion Association [Y2022076].

References

- Al-Lazikani B, Banerji U, Workman P *et al.* Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol* 2012;30:679–92.
- Alqosaibi AI, Abdel-Ghany S, Al-Mulhim F *et al.* Vorinostat enhances the therapeutic potential of Erlotinib via MAPK in lung cancer cells. *Cancer Treatment Res Commun* 2022;30:100509.
- Astashkina A, Mann B, Grainger DW *et al.* A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol Therap* 2012;134:82–106.
- Barretina J, Caponigro G, Stransky N *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603–7.
- Celebi R, Bear Don't Walk O, Movva R *et al.* In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Sci Rep* 2019;9:8949.
- Dalkiran A, Atakan A, Rifaoğlu AS *et al.* Transfer learning for drug-target interaction prediction. *Bioinformatics* 2023;39:i103–10.
- Fan K, Cheng L, Li L *et al.* Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Brief Bioinf* 2021;22:bbab271.
- Fisusi FA, Akala EO. Drug combinations in breast cancer therapy. *Pharm Nanotechnol* 2019;7:3–23.
- Fouquerier J, Guedj M. Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* 2015;3:e00149.
- Ghandi M, Huang FW, Jané-Valbuena J *et al.* Next-generation characterization of the cancer cell line encyclopedia. *Nature* 2019;569:503–8.
- Goswami CP, Cheng L, Alexander PS *et al.* A new drug combinatory effect prediction algorithm on the cancer cell based on gene expression and dose-response curve. *CPT Pharmacom Syst Pharm* 2015;4:80–90.
- He L, Kuleskiy E, Saarela J *et al.* Methods for high-throughput drug combination screening and synergy scoring. In: von Stechow L (ed.), *Cancer Systems Biology: Methods and Protocols, Methods in Molecular Biology*. New York, NY: Springer, 2018, 351–98.
- Hirai H, Sootome H, Nakatsuru Y *et al.* MK-2206, an allosteric Akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs in vitro and in vivo. *Mol Cancer Ther* 2010;9:1956–67.
- Holbeck SL, Camalier R, Crowell JA *et al.* The national cancer institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res* 2017;77:3564–76.

- Hu J, Gao J, Fang X *et al.* DTsyn: a dual-transformer-based neural network to predict synergistic drug combinations. *Brief Bioinform* 2022; 23:bbac302.
- Huang Z, Bianchi F, Yuksekgonul M *et al.* A visual-language foundation model for pathology image analysis using medical Twitter. *Nat Med* 2023;29:2307–16.
- Iqbal MK, Kaur H, Md S *et al.* A technical note on emerging combination approach involved in the onconanotherapeutics. *Drug Deliv* 2022;29:3197–212.
- Janizek JD, Celik S, Lee S *et al.* Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. 2018;33:1769.
- Jiang P, Huang S, Fu Z *et al.* Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Comput Struct Biotechnol J* 2020;18:427–38.
- Koinis F, Kotsakis A, Georgoulias V *et al.* Small cell lung cancer (SCLC): no treatment advances in recent years. *Transl Lung Cancer Res* 2016;5:39–50.
- Kuru HI, Cicek AE, Tastan O *et al.* From cell lines to cancer patients: personalized drug synergy prediction. *Bioinformatics* 2024; 40:btae134.
- Kuru HI, Tastan O, Cicek AE *et al.* MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinf* 2022;19:2334–44.
- Landrum G. RDKit documentation. *Release* 2013;1:4.
- Li J, Huo Y, Wu X *et al.* Essentiality and transcriptome-enriched pathway scores predict drug-combination synergy. *Biology (Basel)* 2020;9:278.
- Li P, Huang C, Fu Y *et al.* Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015;31:2007–16.
- Li R, Tang J, Zheng W J *et al.* SAFER: sub-hypergraph attention-based neural network for predicting effective responses to dose combinations. *Res Sq* 2024.
- Li X, Xu Y, Cui H *et al.* Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif Intell Med* 2017;83:35–43.
- Li Z, Zhu S, Shao B *et al.* DSN-DDI: an accurate and generalized framework for drug-drug interaction prediction by dual-view representation learning. *Brief Bioinform* 2023;24:bbac597.
- Liu J, Gefen O, Ronin I *et al.* Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science* 2020;367:200–4.
- Liu X, Song C, Liu S *et al.* Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction. *Bioinformatics* 2022;38:4782–9.
- Macarron R, Banks MN, Bojanic D *et al.* Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 2011; 10:188–95.
- Malyutina A, Majumder MM, Wang W *et al.* Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Comput Biol* 2019; 15:e1006752.
- Manica M, Oskooei A, Born J *et al.* Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Mol. Pharmaceutics* 2019;16:4797–806.
- O’Neil J *et al.* An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Therap* 2016;15:1155–62.
- Pang Y, Chen Y, Lin M *et al.* MMSyn: a new multimodal deep learning framework for enhanced prediction of synergistic drug combinations. *J Chem Inf Model* 2024;64:3689–705.
- Preuer K, Lewis RPI, Hochreiter S *et al.* DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018; 34:1538–46.
- Ramsundar B, Eastman P, Walters P *et al.* *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More.* California/United States: O'Reilly Media, Inc. 2019.
- Ren S, Yu L, Gao L *et al.* Multidrug representation learning based on pretraining model and molecular graph for drug interaction and combination prediction. *Bioinformatics* 2022;38:4387–94.
- Schneider BJ, Kalemkerian GP. Personalized therapy of small cell lung cancer. In: Ahmad A, Gadgeel SM (eds.), *Lung Cancer and Personalized Medicine: Novel Therapies and Clinical Management, Advances in Experimental Medicine and Biology*. Cham: Springer International Publishing; 2016, 149–74.
- Sicklick JK, Kato S, Okamura R *et al.* Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 2019;25:744–50.
- Sidorov P, Naulaerts S, Ariey-Bonnet J *et al.* Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front Chem* 2019;7:509.
- Sun Z, Huang S, Jiang P *et al.* DTF: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics* 2020;36:4483–9.
- Torkamannia A, Omidi Y, Ferdousi R *et al.* A review of machine learning approaches for drug synergy prediction in cancer. *Brief Bioinform* 2022;23:bbac075.
- Tyers M, Wright GD. Drug combinations: a strategy to extend the life of antibiotics in the 21st century. *Nat Rev Microbiol* 2019; 17:141–55.
- Wang J, Liu X, Shen S *et al.* DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;23:bbab390.
- Wildenhain J, Spitzer M, Dolma S *et al.* Prediction of synergism from chemical-genetic interactions by machine learning. *cels* 2015; 1:383–95.
- Wu Y-Y, Wu H-C, Wu J-E *et al.* The dual PI3K/mTOR inhibitor BEZ235 restricts the growth of lung cancer tumors regardless of EGFR status, as a potent accompanist in combined therapeutic regimens. *J Exp Clin Cancer Res* 2019;38:282.
- Zagidullin B, Aldahdooh J, Zheng S *et al.* DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019; 47:W43–51.
- Zhang H, Wang Z, Nan Y *et al.* Harmonizing across datasets to improve the transferability of drug combination prediction. *Commun Biol* 2023;6:397–10.
- Zhang T, Zhang L, Payne P R O *et al.* Synergistic drug combination prediction by integrating multiomics data in deep learning models. In: Markowitz J (ed.), *Translational Bioinformatics for Therapeutic Development, Methods in Molecular Biology*. New York, NY: Springer US; 2021, 223–38.
- Zheng S, Aldahdooh J, Shadbahr T *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;49:W174–84.