OXFORD

DD-PRiSM: a deep learning framework for decomposition and prediction of synergistic drug combinations

Iljung Jin[1], Songyeon Lee[1], Martin Schmuhalek[2], Hojung Nam (iD)[1,2,3,*]

[1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju 61005, Republic of Korea
[2]AI Graduate School, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju 61005, Republic of Korea
[3]Center for AI-Applied High Efficiency Drug Discovery (AHEDD), Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju 61005, Republic of Korea

*Corresponding author: Hojung Nam, School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju 61005, Republic of Korea. E-mail: hjnam@gist.ac.kr

Abstract

Combination therapies have emerged as a promising approach for treating complex diseases, particularly cancer. However, predicting the efficacy and safety profiles of these therapies remains a significant challenge, primarily because of the complex interactions among drugs and their wide-ranging effects. To address this issue, we introduce DD-PRiSM (Decomposition of Drug-Pair Response into Synergy and Monotherapy effect), a deep-learning pipeline that predicts the effects of combination therapy. DD-PRiSM consists of two predictive models. The first is the Monotherapy model, which predicts parameters of the drug response curve based on drug structure and cell line gene expression. This reconstructed curve is then used to predict cell viability at the given drug dosage. The second is the Combination therapy model, which predicts the efficacy of drug combinations by analyzing individual drug effects and their synergistic interactions with a specific dosage level of individual drugs. The efficacy of DD-PRiSM is demonstrated through its performance metrics, achieving a root mean square error of 0.0854, a Pearson correlation coefficient of 0.9063, and an R^2 of 0.8209 for unseen pairs. Furthermore, DD-PRiSM distinguishes itself by its capability to decompose combination therapy efficacy, successfully identifying synergistic drug pairs. We demonstrated synergistic responses vary across cancer types and identified hub drugs that trigger synergistic effects. Finally, we suggested a promising drug pair through our case study.

Keywords: drug combination response; dose-dependent prediction; synergistic drug combination; cancer treatment; precision medicine

Introduction

The emergence of combination therapies has been a significant advancement in treating complex diseases, particularly cancer and various multifactorial disorders [1, 2]. The rationale for using multiple drugs in one treatment regimen is their potential to target different disease progression pathways, increasing treatment efficacy and potentially reducing drug resistance risk [2–4]. However, this approach adds complexity to treatment planning as interactions between drugs can result in unpredictable therapeutic outcomes and side effects [3, 4]. This unpredictability poses a substantial challenge in the clinical application of combination therapies.

Predicting the efficacy and safety profiles of combination therapies is complex, primarily due to the intricate and often unforeseen interactions among drugs. These interactions, occurring within *in vivo* environments, are influenced by genetic variations, disease heterogeneity, and individual patient responses. Consequently, the traditional trial-and-error method for determining effective drug combinations is time-consuming, costly, and poses significant risks to patient safety.

In response to the growing demand for combination therapy, extensive *in vitro* datasets like NCI-ALMANAC have been established to study combination therapies and their synergistic effects [5]. Computational methods such as DeepSynergy [6] have been developed to predict the synergy of combination therapies by training neural networks. Following DeepSynergy, most subsequent studies about combination therapy like MatchMaker [7], DeepDDS [8], and AttenSyn [9] predicted synergistic metrics based on the Loewe additivity [10] or the Bliss independence [11]. However, these metrics can be misleading as they do not directly estimate the efficacy of the combination but rather calculate its additional effect. For example, a combination might show minimal anticancer efficacy on its own but appear synergistically effective, leading to an overestimated metric. Additionally, those studies did not consider the concentration of each drug, although the dosage combination of drugs is also an important factor for the synergistic effect of the combination therapy. Previous studies focused on the prediction of the representative value like the mean of synergy score, provided by data portals like Drug-Comb [12]. However, those representative values do not reflect the synergistic effect of the combination well. For example, the

combination treatment of 5-fluorouracil and dasatinib is synergistic on the cell line A2058 when treated in 3.25 and 0.024 μM, respectively (Loewe synergy score: 63.22), while it becomes antagonistic with the dosage combination of 0.35 and 3 μM (Loewe synergy score: −12.34); the representative value of Loewe score provided by DrugComb is 10.93.

To address this problem, concentration-level prediction models such as IDACombo [13], comboFM [14], and comboLTR [15] were developed. IDACombo estimates the combination therapy efficacy based on the independent drug action (IDA) model [16], taking the maximum efficacy among monotherapies as the efficacy of the combination therapy. ComboFM predicts the combination therapy efficacy by modeling feature interactions with a factorization machine–based approach. Similarly, ComboLTR also predicts combination therapy efficacy but utilizes latent tensor reconstruction to decompose the tensor product instead [17]. Despite their impressive performance, those studies still have a critical limitation; they can only predict the efficacy of drug combinations when the efficacy scores of individual drugs or identifiers are available. This limitation hinders the practical application of the research as profiles of all cancer patients were not exposed to the model during the training phase.

To overcome these challenges, we introduce the Decomposition of Drug-Pair Response into Synergy and Monotherapy effect (DD-PRiSM), a deep learning pipeline for predicting the effects of combination therapy. DD-PRiSM comprises two predictive models for the prediction of treatment efficacy. The first, the Monotherapy model, is designed to create a drug–response curve by predicting four key parameters that determine the shape of the curve for a specific drug and the corresponding cell lines (Fig. 1A). This model integrates the gene expression, drug structure, and the dosage of the treatment for the prediction of drug response measured as cell viability. The second component, the Combination therapy model, builds upon the cell viability predictions made by the Monotherapy model, aiming to predict the efficacy of drug combinations by analyzing both the individual drug effects and their synergistic interactions (Fig. 1B). We benchmarked the remarkable performance of our model by comparison with state-of-the-art (SOTA) models. After the performance evaluation, we further analyzed synergistic combinations based on the prediction result from DD-PRiSM (Fig. 1C). Finally, we suggested a novel synergistic combination.

Materials and methods
Drug response datasets
Our proposed pipeline, DD-PRiSM, was trained using two datasets: NCI60 [18] and NCI-ALMANAC [5]. First, we pretrained the Monotherapy model using the NCI60 dataset to predict the efficacy of monotherapy treatments. The NCI60 database offers various growth inhibition metrics such as GI50, TGI, and LC50 for each cell line–drug pair after treatment. The model was designed to predict cell viability based on the input of cell line, drug, and its dosage. The NCI60 dataset was obtained from the National Cancer Institute (NCI) Developmental Therapeutics Program (DTP) data portal, specifically the NCI-60 Growth Inhibition Data section. To mitigate differences between the NCI60 and NCI-ALMANAC datasets, we fine-tuned the pretrained Monotherapy model using monotherapy response data available in the NCI-ALMANAC dataset (Table 1). To ensure data quality and compatibility, we performed several preprocessing steps on datasets before training models. For the specification of preprocessing steps, see more details in Dataset Preprocessing and Fig. S1 in Supplementary

Methods. The drugs and cell lines utilized from the NCI-ALMANAC dataset are provided in Supplementary Data 1.

To further assess the model's generalizability, we incorporated an additional combination therapy response dataset from O'Neil *et al.* [19]. Since other studies require tensor indices, we selected combinations involving cell lines, drugs, and concentrations present in the NCI-ALMANAC dataset. This resulted in 1666 combinations composed of seven cell lines and 15 drugs. The "X/X0" value, representing the normalized cell volume relative to the dimethyl sulfoxide (DMSO) control, was used as a cell viability label. It is important to highlight that the O'Neil dataset exhibited a markedly different distribution of input concentrations and combination therapy responses compared to the NCI-ALMANAC (Fig. S2, Supplementary Data 2). The number of cell lines, drugs, drug combinations, and responses are described in Table 1.

Data representations
To represent drugs in our models, we utilized the Simplified Molecular-Input Line-Entry System (SMILES) notation. We directly acquired SMILES representations of compounds in datasets from the NCI DTP Data portal. We used RDKit [20] to convert the SMILES strings into 512-bit Morgan fingerprints, which are binary vectors that encode the presence or absence of specific molecular substructures.

Cell lines were represented with gene expression values grouped by biological pathway. The gene expression data for cell lines were downloaded from the Cancer Cell Line Encyclopedia (CCLE) [21] on the Dependency Map (DepMap) [22] portal. We collected log-transformed transcript per million values of genes in cell lines. For the normalization, we calculated the z-score of gene expressions for each cell line, and the sum of all normalized gene expressions in a cell line became zero.

For the biological network of the Monotherapy model, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [23] was obtained from the Molecular Signatures Database (MSigDB) [24, 25]. The latest version of the KEGG pathway gene set (KEGG-legacy), which consists of 186 pathways, was used for this study.

Monotherapy response prediction
To calculate cell viability at specific drug dosages, we utilized the response curve of cancer–drug interactions, which is commonly modeled using a four-parameter sigmoid function based on the Hill equation [26]:

$$y = y_{min} + \frac{y_{max} - y_{min}}{1 + e^{k(x - IC50)}}$$

where y_{min} represents the minimum viability, y_{max} is the maximum viability, $IC50$ is the half maximal inhibitory concentration, and k defines the steepness of the response curve around the IC50 concentration. The Monotherapy model predicts those four parameters of the sigmoid function, building upon modifications from our earlier study, Hierarchical Network for Drug Response Prediction with Attention (HiDRA) [27]. By predicting these parameters, the model is able to reconstruct the entire dose–response curve, which describes the relationship between drug dosage and cell viability for each cell line–drug pair. Given a specific dosage, we input this value into the reconstructed curve to estimate the resulting cell viability following monotherapy treatment at that dosage. This approach allows for a more flexible and informative prediction of cell viability compared to predicting a single value at a fixed concentration. In addition to predicting cell viability, the Monotherapy model predicts a metric termed pathway attention
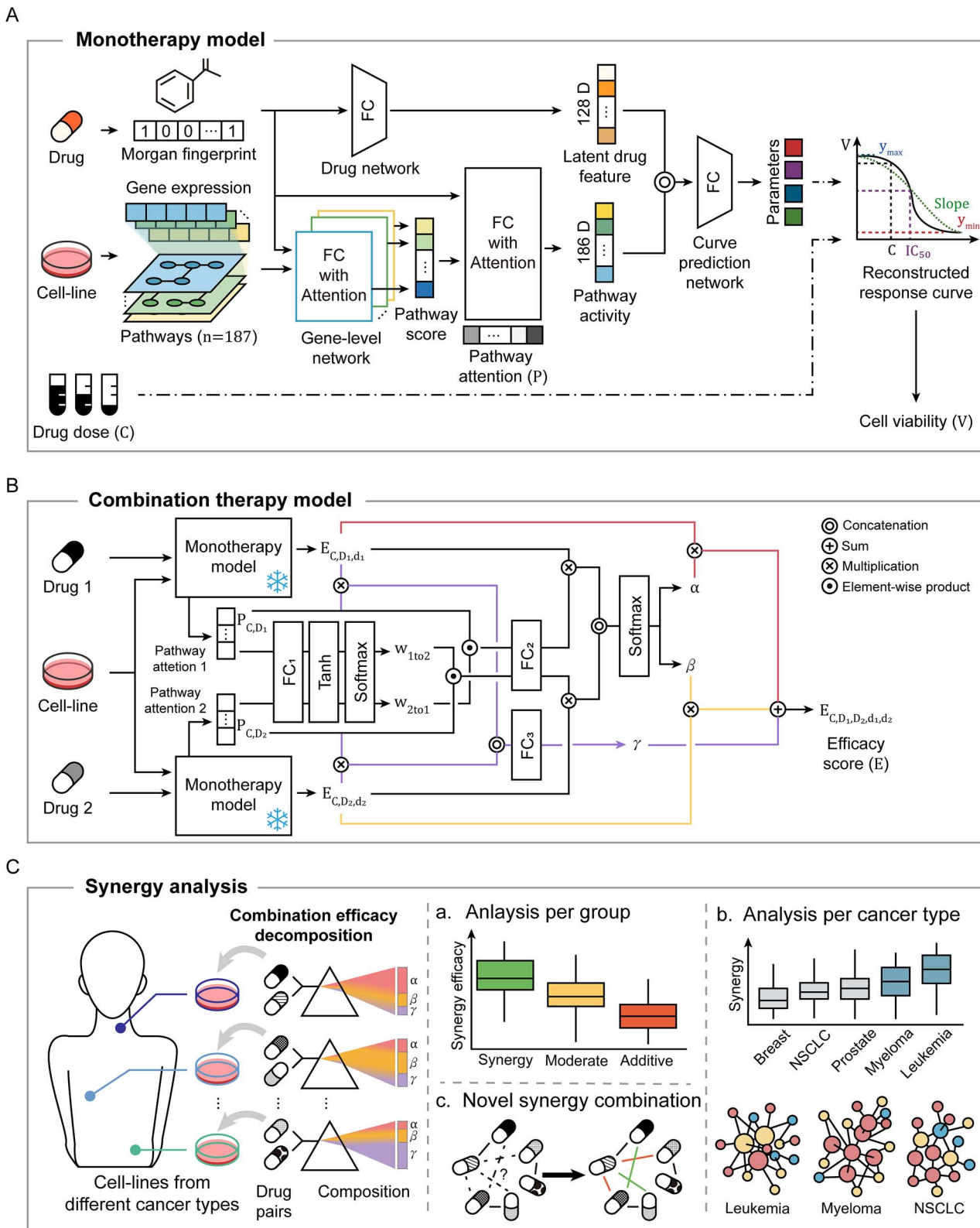
Figure 1. Schematic overview of DD-Prism. The study consists of two predictive models and an analytical component. (A) The Monotherapy model. This model reconstructs drug-response curves by learning four key parameters that shape the curves for specific drugs and their corresponding cell lines. (B) The Combination therapy model. Building upon the predicted cell viability data derived from the Monotherapy model, this model operates with drug pairs and corresponding cell lines to predict the efficacy of drug combinations by analyzing both individual drug effects and their synergistic interactions. (C) Analysis of prediction results. This component highlights the application of the model's predictive outcomes in analyzing drug interactions.

Table 1. Details of the preprocessed datasets used in this study.

Dataset	Number of cell lines	Number of drugs	Number of drug combinations	Number of responses
NCI60 (monotherapy pretraining)	66	50 893	Not available	10 105 780
NCI-ALMANAC (monotherapy fine-tuning)	44	102	Not available	35 041
NCI-ALMANAC (combination therapy training)	44	102	5032	1 981 135
O'Neil dataset (external validation)	7	15	77	1666

(P), which represents the importance of various biological pathways in the drug-treated cell line.

The specific architecture of the Monotherapy model, including the input features, hidden layers, and pathway attention mechanism, is detailed in our previous study, HiDRA.

Combination therapy response prediction

The Combination therapy model predicts the efficacy of drug combinations based on the hypothesis that drug interactions can be deduced from each drug's mechanism of action on the cell line, represented as pathway attention in our study. The efficacy of drug combination is modeled using the following formula:

$$V = 1 - E$$

where:
  V is cell viability.
  E is treatment efficacy.

$$E_{C,D_1,D_2,d_1,d_2} = \alpha E_{C,D_1,d_1} + \beta E_{C,D_2,d_2} + \gamma$$

where:
  C represents the cell line.
  D_1 and D_2 are the two drugs in the combination.
  d_1 and d_2 are the respective dosages of the drugs.
  E_{C,D_1,d_1} and E_{C,D_2,d_2} represent monotherapy efficacies.
  α and β are weights for the monotherapy efficacies.
  γ represents the synergistic effect.

The response to the drug combination can be decomposed into three components: (i) the monotherapy effect of Drug 1, (ii) the monotherapy effect of Drug 2, and (iii) the synergy effect.

The weights for monotherapy effects determine how much each monotherapy affects the combination therapy. Two weights, α for Monotherapy 1 and β for Monotherapy 2, are predicted by considering the pathway attention and the efficacy of two monotherapies using fully connected (FC) neural networks.

$$w_{2\text{to}1} = \text{softmax}\left(\text{Tanh}\left(\text{FC}_1\left(P_{C,D_2}\right)\right)\right)$$

$$\alpha_{\text{pre}} = E_{C,D_1,d_1} * \text{FC}_2\left(P_{C,D_1} \odot w_{2\text{to}1}\right)$$

$$w_{1\text{to}2} = \text{softmax}\left(\text{Tanh}\left(\text{FC}_1\left(P_{C,D_1}\right)\right)\right)$$

$$\beta_{\text{pre}} = E_{C,D_2,d_2} * \text{FC}_2\left(P_{C,D_2} \odot w_{1\text{to}2}\right)$$

$$\alpha, \beta = \text{softmax}\left(\left[\alpha_{\text{pre}} \| \beta_{\text{pre}}\right]\right)$$

Here, P_{C,D_1} and P_{C,D_2} represent the pathway attention from Monotherapy 1 and Monotherapy 2, $\odot$ denotes the Hadamard product, and $[\alpha \| \beta]$ represents the concatenation of two vectors.

γ, the synergy effect of two monotherapies, is also calculated with the same inputs: two efficacies and two pathway attentions from monotherapies.

$$\gamma = \text{FC}_3\left(\left[E_{C,D_1,d_1} * P_{C,D_1} \| E_{C,D_2,d_2} * P_{C,D_2}\right]\right)$$

The architecture of the FC neural network FC_3 is identical to FC_2, but they don't share network parameters.

For the specification of the model, number of layers, and number of nodes, see more details in Table S1 in Model Specification in Supplementary Methods.

Training scheme

For the training and evaluation of the Monotherapy model, we divided each monotherapy dataset into two sets: a training set (90% of the data) for model training and validation and a test set (10% of the data). The test set comprised four subsets: the unseen pair set, the unseen cell line set, the unseen drug set, and the unseen all set. To create this test set, we first sampled cell lines and drugs to serve as unseen entities, aiming to make each subset ∼2.5% of the total dataset. Pairs with either unseen cell lines or unseen drugs were categorized into stratified test sets: the unseen cell line set and the unseen drug set, respectively. Pairs containing both an unseen cell line and an unseen drug were grouped into the unseen all set. From the remaining pairs with seen cell lines and drugs, 5% were allocated to the unseen pair set. The remaining samples were used as a training–validation set, which was further split into a training set and a validation set in an 8:1 ratio (Tables S2 and S3). Each pair had three to five concentrations with corresponding cell viability measurements, and all dosage-dependent data were utilized during the model training and testing phases.

As detailed in the Drug Response Datasets section, our Monotherapy model training utilized two datasets: NCI60 and NCI-ALMANAC. The first dataset, NCI60, is a large monotherapy dataset with a sufficient number of samples to effectively train interactions between cancer cells and drugs. However, since it lacks a combination therapy response, it is not suitable to train the entire DD-PRiSM framework. The second dataset, NCI-ALMANAC, contains both monotherapy response data and corresponding combination therapy data but has fewer monotherapy samples. We employed a two-step training process to leverage the strengths of both datasets. First, we used the extensive NCI60 dataset to pretrain the Monotherapy model (Fig. S3A). This step allowed the model to learn broad cancer–drug interactions from a diverse range of compounds and cell lines. Second, we fine-tuned the pretrained model trained with the training set of NCI60 using the monotherapy portion of the NCI-ALMANAC dataset (Fig. S3B). This step aimed to adapt the model to the specific characteristics of the NCI-ALMANAC data, which is directly relevant to our combination therapy predictions. During fine-tuning, we froze most model parameters to preserve the model's ability to infer

cancer–drug interactions through biological pathways. Only the final layers that predict the four parameters of the response curve were unfrozen, allowing adjustment for batch effects between datasets.

For the Combination therapy model, we utilized the combination data from the NCI-ALMANAC dataset for model training and evaluation. The test set was constructed with unseen cell lines and drugs in alignment with the NCI-ALMANAC monotherapy dataset. Specifically, we created two drug-related unseen datasets: the unseen one drug set (where only one drug was excluded from the training and validation sets) and the unseen two drug set (where both drugs were excluded) (Table S4). The details of the training scheme are described in Detailed Training Scheme in Supplementary Methods.

In the training process, we integrated both error-based and correlation-based loss functions. We used mean squared error (MSE) as the primary error-based loss function. To address the tendency of error-based loss functions to converge toward the training set average, we incorporated a correlation-based loss function. Inspired by Atmaja and Akagi [28], we implemented a correlation-harmonized loss function $\mathcal{L}_{corr}$.

$$\mathcal{L}_{corr} = 1 - \frac{\sum (y - \bar{y})(\hat{y} - \bar{\hat{y}})}{\sqrt{\sum (y - \bar{y})^2}\sqrt{\sum (\hat{y} - \bar{\hat{y}})^2}}$$

where y is a true label, $\hat{y}$ is a predicted label, and $\bar{y}$ represents the average of y values.

To address the class imbalance in cell viability, we implemented sample density weighting in the loss function. We estimated the density of cell viability distribution using a Gaussian kernel and divided the training dataset's cell viabilities into fixed-width bins (width = 0.1). Each sample's cell viability was then assigned a density-based weight, which was applied to the error loss to account for this imbalance. The density weighted MSE $\mathcal{L}_{dMSE}$ is calculated as follows:

$$\mathcal{L}_{dMSE} = \frac{1}{n} \sum (1 - \gamma d)(y - \hat{y})^2$$

where d is the density of a given sample's cell viability on the sample distribution estimated on the training dataset, and γ is a hyperparameter controlling the strength of density-based weighting. All training steps in this study were trained with an aggregated loss function $\mathcal{L}_{total}$.

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{dMSE} + \beta \mathcal{L}_{corr}$$

For our study, we set loss-related hyperparameters as $\alpha = 1$, $\beta = 0.5$, and $\gamma = 0.75$. We used the AdamW optimizer [29] to update model parameters.

To optimize the training process, we implemented learning rate reduction and early stopping. We initialized the learning rate at 1e-2 for all training steps except the fine-tuning (1e-3). The learning rate was reduced by a factor of 10 if the validation loss did not decrease by more than a threshold value (0.0005) for 10 consecutive epochs. Similarly, we employed early stopping to prevent overfitting, terminating training if the validation loss did not decrease by more than the threshold (0.0005) for 20 consecutive epochs.

Performance evaluation

To evaluate the performance of the proposed work, we used four metrics widely used for the regression task: root mean square error (RMSE), Pearson correlation coefficient (PCC), coefficient of determination (R^2), and concordance index (C-index). We further evaluated our framework by comparing each module with previously published studies. For the Monotherapy model, we compared its performance against support vector regression (SVR), a baseline machine learning model, and DeepTTA [30], a transformer-based model for predicting monotherapy response. Since those models predict IC50 value, a dosage-independent drug response metric, we trained and tested them with IC50 values derived from the NCI-ALMANAC monotherapy data's cancer–drug response curves. These curves were reconstructed using a curve fitting function from the SciPy package [31]. To ensure robust evaluation, we generated 10 different training, validation, and test sets by splitting the NCI-ALMANAC monotherapy dataset using 10 distinct random seeds. For each seed, the training set, the validation set, and the test set were in an 8:1:1 ratio. Our Monotherapy model was trained on cell viability data corresponding to the IC50 training and validation sets, and IC50 values for the test sets were obtained from the reconstructed response curves corresponding to the IC50 test set.

The Combination therapy model by comparing its performance against two previously published dosage-dependent combination response prediction studies: comboFM [14], comboLTR [15], and IDACombo [13]. ComboFM predicts cell viability after combination therapy using a factorization machine approach. It incorporates molecular fingerprints, gene expression data, and entity identifiers as input features. For our comparison, we implemented a simplified version of comboFM, called comboFM-2, due to limitations in available packages. Specifically, we developed a factorization machine of degree two, where the output is calculated as the weighted sum of features along with the weighted sum of the pairwise interactions between two features. The comboFM-2 model was implemented using the PyTorch framework [32]. Another combination therapy efficacy prediction model, comboLTR, employs the latent tensor reconstruction [17], a tensor decomposition approach similar to the factorization machine. For comboLTR, we used tensor indices and Molecular ACCess System (MACCS) fingerprints, as omics data such as CRISPR-Cas9 profiles were unavailable for most cell lines. It is important to highlight that the authors of comboLTR reported only a modest performance improvement when incorporating cell-line omics data into tensor indices and the MACCS fingerprints. IDACombo is a study based on the IDA model [16]; this method computes the efficacy of combination therapy by selecting the larger of the two individual monotherapy efficacies. To ensure a fair comparison with IDACombo, we used ground truth monotherapy values instead of those predicted by our Monotherapy model, as IDACombo also requires and uses these values. This approach allows for a more direct comparison of the combination prediction capabilities of each model. We generated 10 datasets, each consisting of a training–validation set and a test set, by splitting the NCI-ALMANAC combination therapy data with 10 different random seeds. For each collection, the training–validation set and test set were split in a 9:1 ratio. The training–validation set was further divided into a training set and a validation set with an 8:1 ratio.

Beyond internal validation, we evaluated the performance of DD-PRiSM against other deep learning–based models using the O'Neil combination therapy response dataset to demonstrate the broader generalizability of DD-PRiSM compared to other

approaches. IDACombo was excluded from this external validation due to its lack of trainable parameters.

The details of the performance comparison and dataset statistics for the performance comparison are specified in Performance Comparison with Related Studies in Supplementary Methods.

Results
Model performance evaluation

We first investigated the performance of the pretrained Monotherapy model. The model exhibited outstanding performance on the unseen pair set of the NCI60 dataset, achieving an RMSE of 0.0830, a PCC of 0.9387, and an R^2 of 0.8811 (Fig. 2A). To assess the generalizability of the Monotherapy model, we evaluated the performance of it on three stratified test sets: unseen cell line set, unseen drug set, and unseen all set. The performance of the pretrained Monotherapy model on stratified test sets is presented in Fig. S4.

Following pretraining on the NCI60 dataset, we fine-tuned the Monotherapy model using NCI-ALMANAC data points. Despite limited data availability, our Monotherapy model demonstrated strong predictive capabilities on the NCI-ALMANAC dataset. Notably, it showed an RMSE of 0.0914, a PCC of 0.8791, and an R^2 of 0.7725 on the unseen pair set (Fig. 2B). The fine-tuned Monotherapy model was similarly evaluated on the stratified test sets, as was done with the pretrained model. The performance results on these test sets are shown in Fig. S5. These results highlight the model's effective generalization to the new NCI-ALMANAC dataset.

Next, we investigated the performance of the Combination therapy model on the combination therapy response data of NCI-ALMANAC. The Combination therapy model achieved impressive performance, with an RMSE of 0.0854, a PCC of 0.9063, and an R^2 of 0.8209, on the unseen pair set (Fig. 2C). However, its performance was relatively lower on the stratified test sets, with a PCC of ~0.75 across all sets. The Combination therapy model's performance on these stratified test sets is illustrated in Fig. S6. Notably, through our data ablation study (Fig. S7), we found that both pretraining and fine-tuning steps of the Monotherapy model were crucial for the Combination model's effectiveness—omitting either step led to overfitting and significantly reduced generalizability. This highlights the importance of our two-stage training approach in achieving robust predictive performance.

We further examined model performance across error intervals (Fig. 2D). Approximately 90% of the test set showed a correlation of 0.9 or higher, while the remaining 10% had a PCC of ~0.625. This demonstrates DD-PRiSM's generalizability, accurately predicting cell viability for most test data points. Next, we assessed model performance for each cancer type (Fig. 2E). Despite slight variations among cancer types, the model performed well across all categories, with PCC values ~0.9 and RMSE values ~0.1.

Performance comparison with related studies

To evaluate the robustness of the DD-PRiSM framework, we compared its performance against other models addressing similar challenges. First, we compared the Monotherapy model with SVR, a baseline machine learning model, and DeepTTA, the transformer-based SOTA model. For this analysis, we used the monotherapy responses from the NCI-ALMANAC, which was used as the fine-tuning dataset for the Monotherapy model (Table S5). Since the compared models predict IC50 directly, cell viability values of our model were converted to IC50 values through curve fitting based on the Hill equation. As a result, the Monotherapy

model outperformed SVR but performed slightly below DeepTTA (Fig. 3). Notably, our model achieved competitive performance (PCC of 0.8906, averaged over 10 iterations) compared to DeepTTA (PCC of 0.9219) despite facing an additional challenge. While DeepTTA was specifically trained to predict IC50 values directly, our Monotherapy model was trained to predict cell viability values, from which we then had to extract IC50 values for this comparison. A key advantage of our Monotherapy model is its ability to predict drug efficacy across any concentration, as it is dosage-dependent. In contrast, DeepTTA is limited to predicting the dosage-independent IC50 value, which does not provide information on how effective a drug will be at specific dosages. This makes our model more practical for clinical applications as it can directly assess whether a particular dosage will effectively treat cancer—a critical question that DeepTTA cannot address.

Next, we compared the performance of the Combination Therapy model with other dosage-dependent studies, including two deep learning–based approaches, comboFM and comboLTR, as well as the rule-based model, IDACombo on combination therapy responses from the NCI-ALMANAC (Table S6). It should be noted that the comboFM model was implemented as a second-order factorization machine that the authors named comboFM-2, and the comboLTR model was implemented without inaccessible cell-line multi-omics features, referred to by the authors as "tensor indices + MACCS."

Compared with deep learning–based studies, DD-PRiSM demonstrated a modest performance improvement with statistical significance across most metrics (Fig. 4A, Fig. S8). However, DD-PRiSM was slightly outperformed by comboFM in terms of the C-index, indicating that DD-PRiSM provides more precise predictions of the actual response values, while comboFM is more accurate in predicting the rank order of responses. As for comboLTR, DD-PRiSM outperformed it across all metrics.

Meanwhile, DD-PRiSM showed performance improvement with statistical significance compared to the IDACombo (Fig. 4B, Fig. S9). Specifically, DD-PRiSM showed an 8.37% improvement in RMSE, a 1.54% increase in PCC, and a 3.07% improvement in R^2 compared to IDACombo.

A notable advantage of DD-PRiSM over previous dosage-dependent studies is its ability to predict combination therapies for unseen cell lines or drugs (Fig. S6). In contrast, earlier models face significant limitations when dealing with unseen entities. For instance, comboFM and comboLTR rely on one-hot vectors to represent cell lines, drugs, and concentrations. These identifier vectors play a crucial role in the model, as comboFM is based on a factorization machine. Factorization machines excel at capturing relationships between features and perform well in recommendation predictions using one-hot vectors. However, this approach inherently limits the model's ability to generalize to unseen entities. Similarly, IDACombo has its own constraints. Based on the IDA theory, it requires response values for each monotherapy treatment. This dependency on pre-existing monotherapy data restricts its applicability to new drug–cell line combinations. Unlike these previous studies, DD-PRiSM does not face such constraints when dealing with unseen data. This flexibility allows DD-PRiSM to make predictions for novel cell lines and drugs, significantly expanding its potential applications in combination therapy research.

Performance evaluation on an external dataset

To assess the robustness of DD-PRiSM, we evaluated it alongside other models using the O'Neil dataset, selecting the model with the lowest loss from NCI-ALMANAC training.
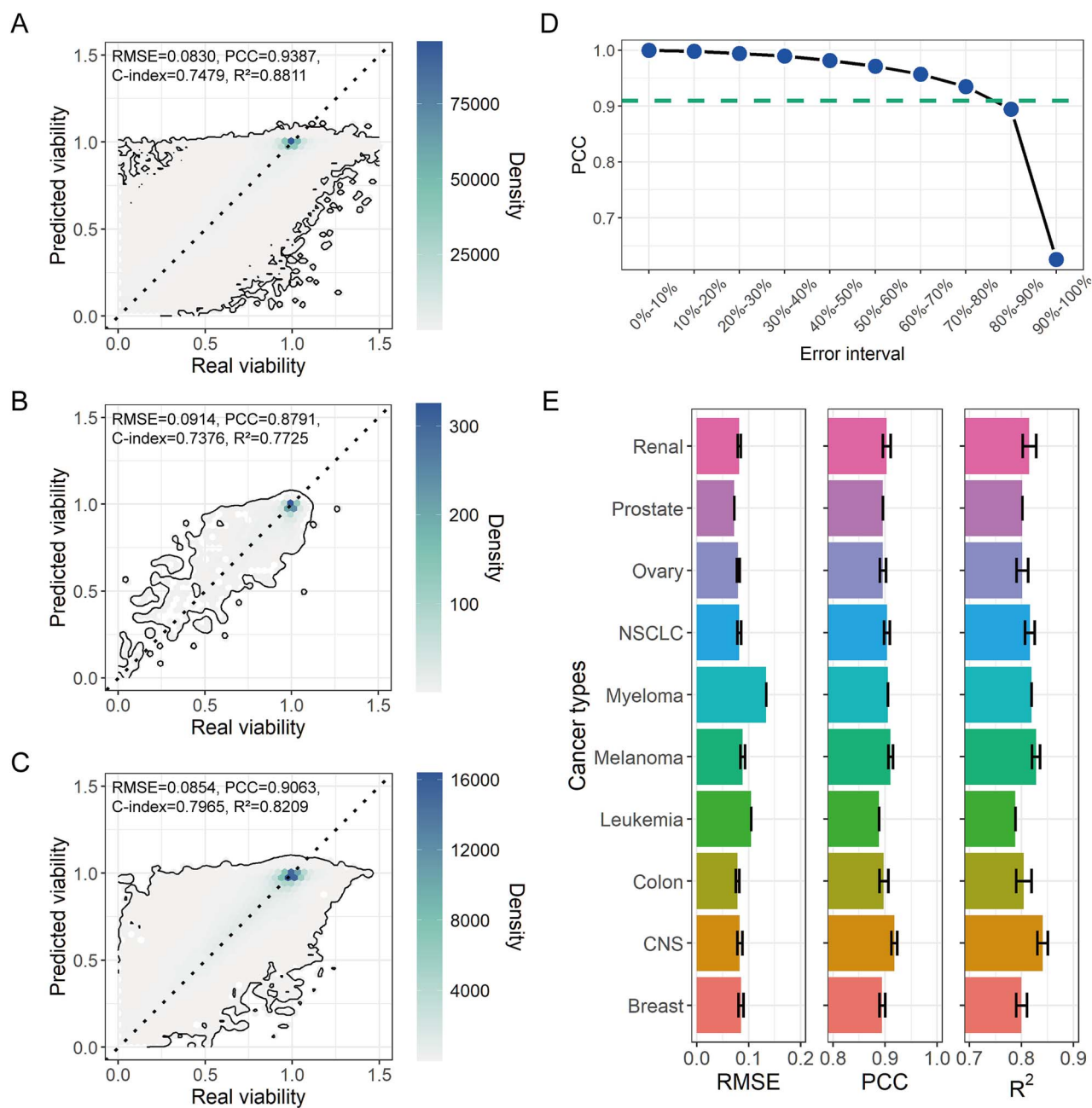
Figure 2. Performance of DD-PRiSM. (A) Pretrained Monotherapy model performance on NCI60 dataset. (B) Fine-tuned Monotherapy model performance on NCI-ALMANAC monotherapy response data. (C) Combination therapy model performance on NCI-ALMANAC combination therapy response data. (D) PCC between ground truth responses and predictions for the combination therapy dataset, grouped by error interval. (E) Combination therapy model performance across different cancer types. All panels display results from the unseen pair set.

As a result, among the compared models, DD-PRiSM achieved the highest performance with a C-index of 0.5968, which was 11.87% higher than comboFM and 17.67% higher than comboLTR (Fig. 5). However, all models exhibited reduced performance on the O'Neil dataset due to its out-of-distribution nature compared to NCI-ALMANAC, especially for the R^2 values.

It is well known that the performance of data-driven models, such as deep learning, significantly drops when applied to out-of-distribution data [33]. In our case, the O'Neil dataset differs substantially from NCI-ALMANAC in both drug concentration distribution and efficacy. Most notably, the O'Neil dataset features higher drug concentrations (Fig. S2A, Supplementary Data 2), resulting in lower cell viability measurements compared to NCI-ALMANAC (Fig. S2B). This distribution shift poses a particular challenge since all models, including DD-PRiSM, were trained on NCI-ALMANAC data where combinations typically show higher viability. Furthermore, models that strongly rely on tensor indices, like comboFM or comboLTR, are likely to exhibit significant performance drops on out-of-distribution data, as the combinations of tensor indices differ from those seen during training steps.

Large-scale evaluation of synergistic effects in drug combinations

Following the performance evaluation, we conducted a comprehensive analysis of treatment efficacy across all possible cell lines and drug combinations presented in the NCI-ALMANAC dataset, using the concentrations provided in the NCI-ALMANAC dataset,
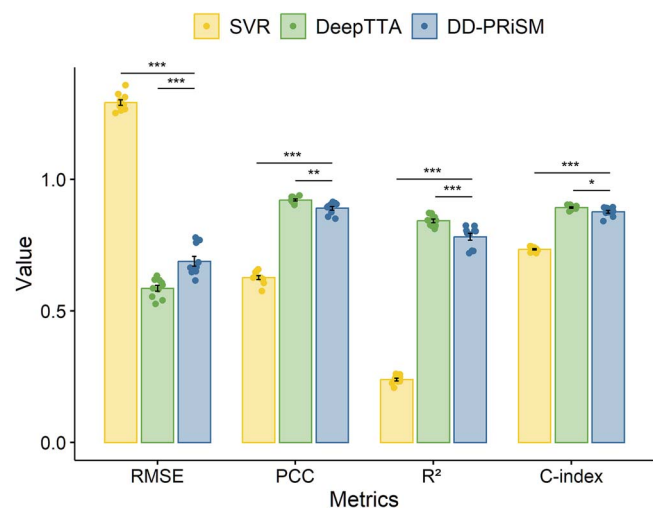
Figure 3. The performance of DD-PRiSM compared with previous dosage-independent monotherapy response prediction models. Asterisks show the statistical significance from the Mann–Whitney U test; * ($P < .001$).

as the effects of monotherapies and the synergistic effects hypothesized by our model vary depending on the dosage of each monotherapy (Fig. 6A). Our primary objectives were to identify the characteristics of synergistic combinations and discover novel synergistic pairs that could potentially lead to more effective cancer treatments.

We analyzed the treatment efficacies of drug combination results on an extensive dataset comprising 81 million possible pairs of data points with varying concentrations. This large-scale analysis revealed a crucial insight: the majority of the treatment efficacy in drug combinations is predominantly driven by the effects of individual drugs rather than synergistic interactions. This conclusion is substantiated by examining the distribution of the γ value, a key indicator of synergy. We observed that the highest peak of the γ value distribution is significantly biased toward 0, which indicates no synergistic effect (Fig. 6B). This finding suggests that true synergistic interactions, while potentially very impactful, are relatively rare among all possible drug combinations, as reported in previous studies [34, 35].

To focus our investigation on identifying drug combinations that exhibit synergism, we developed a more refined approach. We introduced a "reachable synergy" score for each pair of a cell line and drugs. This score was calculated by taking the maximum synergy value across different concentrations, allowing us to capture the highest potential for synergy in each combination, regardless of the specific concentration at which it occurred. Using this reachable synergy score, we then applied a clustering technique to analyze patterns across the entire dataset. Specifically, we employed the SpectralBiclustering algorithm [36] from the scikit-learn package [37]. We applied this clustering method to the reachable synergy values of 226 644 combinations, which represented all possible pairings of 44 cell lines with 5151 drug combinations (Fig. 6C). The clustering analysis yielded three distinct clusters of drug combinations, each with unique characteristics. First, the synergistic group with 1739 combinations. These represent drug pairs that consistently demonstrated high levels of synergy across multiple cell lines. Second, the moderate group encompassed 1977 combinations. These drug pairs showed some synergy but not as consistently or strongly as those in the synergistic group. Third, the additive group includes 1435 combinations. These drug pairs primarily exhibited additive effects,

meaning their combined efficacy was approximately the sum of their individual effects without a significant synergistic boost. Importantly, these three groups showed statistically significant differences in the distribution of their reachable synergy scores (Fig. 6D).

Next, we investigated the therapeutic categories for each cluster and examined the proportion of each category within the groups (Fig. 6E). This analysis revealed intriguing patterns in drug combination efficacy. The synergistic group displayed a significantly higher proportion of combinations that included targeted therapies, as also shown in Jaaks *et al.* [34]. This finding suggests that the interplay between traditional cytotoxic agents and molecularly targeted drugs often leads to enhanced therapeutic effects, possibly due to their complementary mechanisms of action. Conversely, combinations involving non-anticancer drugs were enriched in the additive group. This observation is particularly interesting as it implies that drugs not primarily designed for cancer treatment tend to have less synergistic potential when combined with anticancer agents. However, their presence in combination therapies might still be valuable for managing side effects or improving overall treatment outcomes through additive effects.

Synergism by the cancer type

Previous studies have reported that the synergy effect of combination therapy varies depending on the patient and the cancer type [34, 38]. We investigated the reachable synergy score for each cancer type and compared the distribution of them. To analyze intensive and informative synergistic combinations, we filtered data points with two conditions: concentration and sensitivity. First, we selected the data point whose concentration range was close to the original NCI-ALMANAC. Second, we used data points whose combination therapy efficacy is sensitive, but both monotherapies were not. For this process, we considered treatment as sensitive if the cell viability after the treatment was smaller than 30%.

Interestingly, blood cancer cell lines showed significantly higher reachable synergy than solid tumor cell lines (Fig. 7A). This result was consistent with previous findings that hematologic malignancies get synergistic benefits from combination therapy easily, while combination therapy–based approaches for solid neoplasms are not well developed [39].

Next, we constructed the drug synergism network with synergistic combinations for each cancer type. We chose the top 30% synergistic combinations for each cancer type and made a network that showed the synergistic characteristics of each drug. First, we analyzed the drug synergism network of leukemia (Fig. 7B), the most synergistic cancer type. We found that three drugs act as synergistic agents for leukemia: dasatinib, azacitidine, and bortezomib. A literature survey reveals that dasatinib, a hub drug with the highest degree, is a promising agent for combination therapy in leukemia [40, 41]. Combination therapies with the second hub drug, azacitidine, were also supported by clinical trials: NCT01038635 (azacitidine + lenalidomide; 21st from 234 combinations in our leukemia network) [42], NCT00948064 (azacitidine + vorinostat; 38th from 234 combinations in leukemia network) [43]. The last hub drug, bortezomib, was also supported by literature [44].

Then, we focused on lung cancer with the cancer-type non-small cell lung cancer (NSCLC) to construct the drug synergism network (Fig. 7C). In contrast to the case of leukemia, the drug synergism network of NSCLC did not reveal any dominant hub drug. A drug with the highest degree in the network was pemetrexed, a
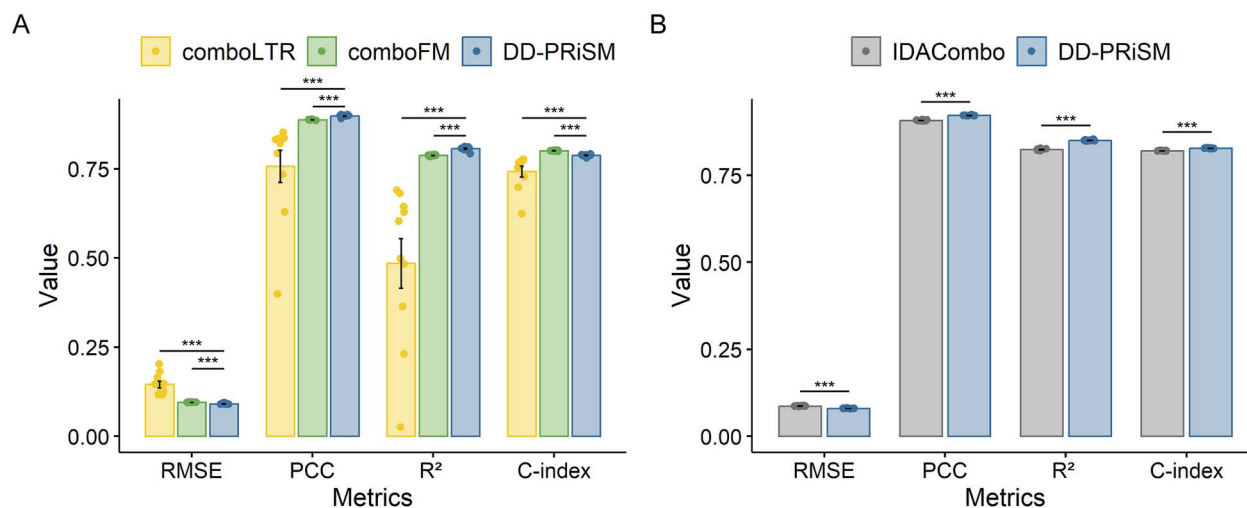
Figure 4. The performance of DD-PRiSM compared with previous dosage-dependent combination therapy response prediction studies. (A) Performance comparison with deep learning–based studies. (B) Performance comparison with IDACombo, the rule-based study. As IDACombo uses the ground truth monotherapy viability, we also used the ground truth monotherapy for DD-PRiSM in this comparison. Asterisks showed the statistical significance from the Mann–Whitney U test; * ($P < .001$).
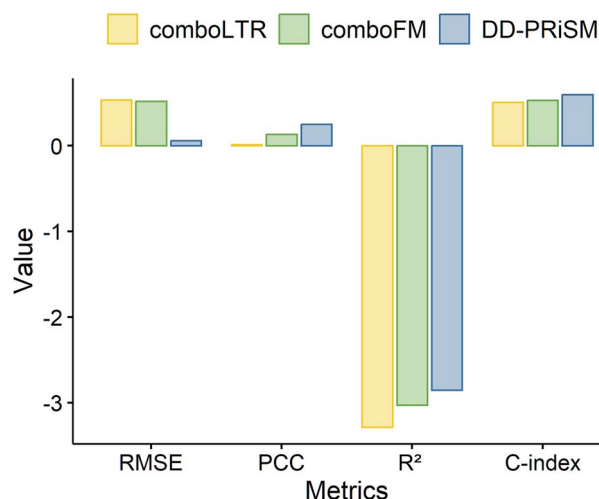

Figure 5. The performance of DD-PRiSM and previous dosage-dependent studies on the external validation. For each compared study, the value in the parenthesis is the performance improvement of DD-PRiSM compared to the study.

widely adopted antifolate agent. We found several literatures that investigated the synergistic potentiality of pemetrexed [45, 46]. Additionally, there were many clinical trials that supported the synergistic effect of pemetrexed in combination with platinum-based antineoplastic agents [47, 48]. Carboplatin, the second-highest-degree drug in our NSCLC network, is also mentioned as a synergistic agent with pemetrexed, and the combination of pemetrexed and carboplatin had a high reachable synergy score in our network (65th from 340 combinations in the NSCLC network) [48, 49].

Synergistic combinations for each cancer type are specified in Supplementary Data 3.

Prediction case study

Based on our previous findings, we conducted an in-depth examination of combinations within the synergistic group that were not present in the model training dataset. This approach allowed us to assess the model's predictive capabilities on novel drug pairs and potentially uncover clinically relevant synergistic combinations.

Our case study focused on the combination of a proteasome inhibitor bortezomib and a DNA methyltransferase inhibitor azacitidine (Fig. 8A). Neither of these drugs was included in the training data, presenting an ideal opportunity to test the model's generalization ability. Notably, existing literature has demonstrated the synergistic effects of this combination against multiple myeloma cells [50, 51], providing a valuable benchmark for our predictions. We concentrated our analysis on the RPMI-8226 cell line, a well-established model for multiple myeloma. To begin, we evaluated the performance of our Combination therapy model on this specific drug pair. The model exhibited exceptional predictive accuracy, with an RMSE of 0.1450, a PCC of 0.9378, and an R^2 of 0.8751 (Fig. 8B). These robust performance metrics lend substantial credibility to the subsequent detailed analyses and interpretations of the model's predictions.

Following this validation, we investigated decomposed synergy values predicted by our model (Fig. 8C and D). The combination of RPMI-8226 + bortezomib + azacitidine revealed intriguing patterns. We observed that the overall magnitude of the synergistic effect escalated in correlation with increasing concentrations of azacitidine. This dose-dependent relationship suggests that higher doses of azacitidine may be crucial in maximizing the synergistic potential of this combination.

Interestingly, our model's predictions for bortezomib revealed a more complex pattern. High synergy efficacy was predicted not only at elevated concentrations but also at lower doses (Fig. 8D). While the magnitude of the synergistic effect varied with bortezomib concentration, our results strongly indicated that the mere presence of bortezomib, rather than its concentration, was the primary driver of synergy.

Discussion

In this study, we introduced DD-PRiSM, a novel framework for predicting the efficacy of combination therapy using predicted monotherapy efficacy and the effects of monotherapy on biological pathways. DD-PRiSM addresses several key challenges in the field of drug combination prediction, offering significant advantages over previous approaches.

A major strength of DD-PRiSM is its ability to predict anticancer therapy response in a concentration-dependent manner, which
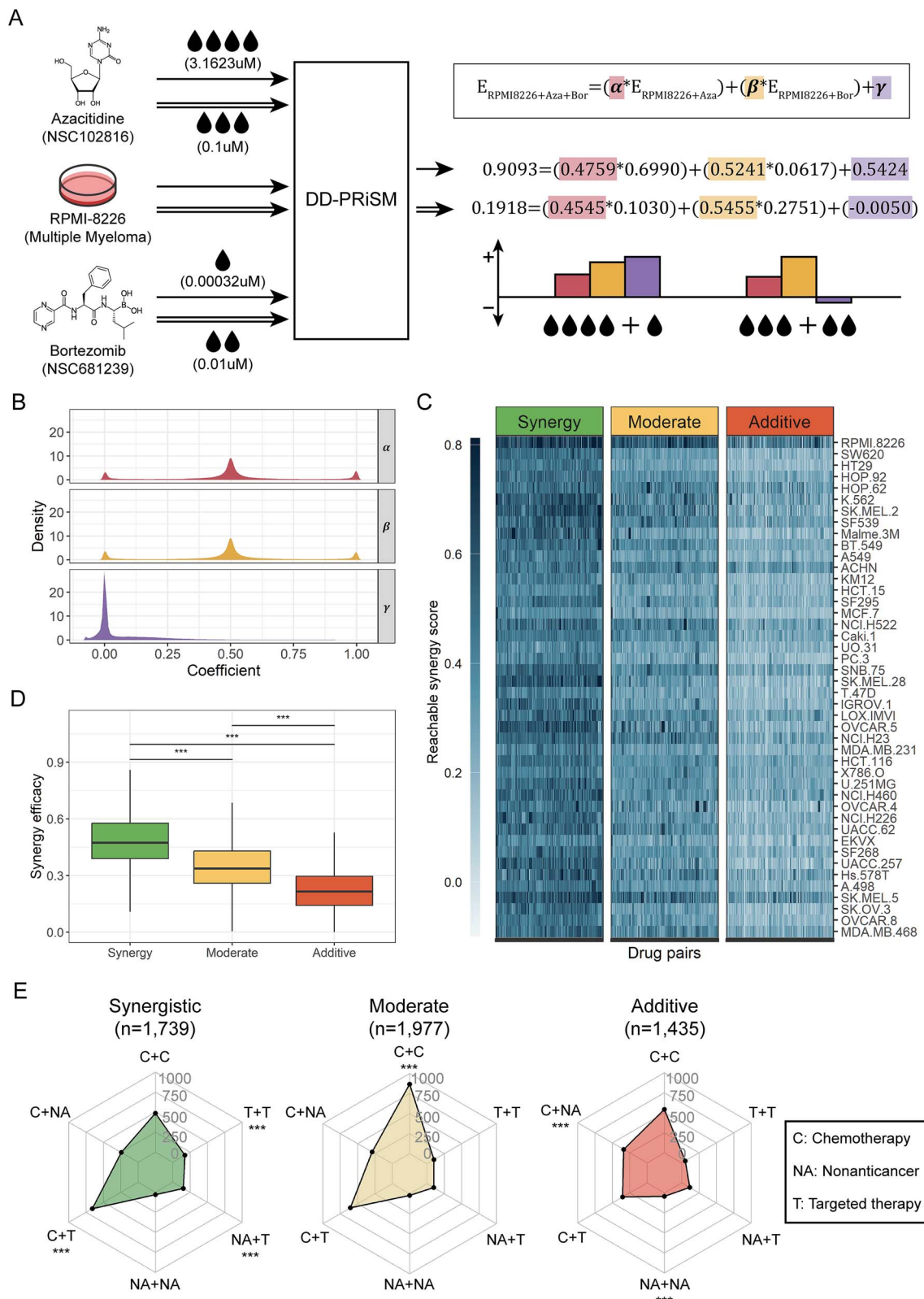
Figure 6. Drug combination analysis of DD-PRiSM's results. (A) Synergy effects differ across the concentration of the drug combination. (B) Distribution of three coefficients, α: the coefficient of the Monotherapy 1, β: the coefficient of the Monotherapy 2, and γ: the synergy effect, for all valid combinations. (C) The clustering results on the reachable synergy from all valid combinations. The left group is a synergistic group, the middle group is a moderate group, and the right group is an additive group. (D) The distribution of reachable synergy for each group. (E) The composition of combinations for each group. For all panels, asterisks showed the statistical significance of the hypergeometric test with the group; * (P < .001).
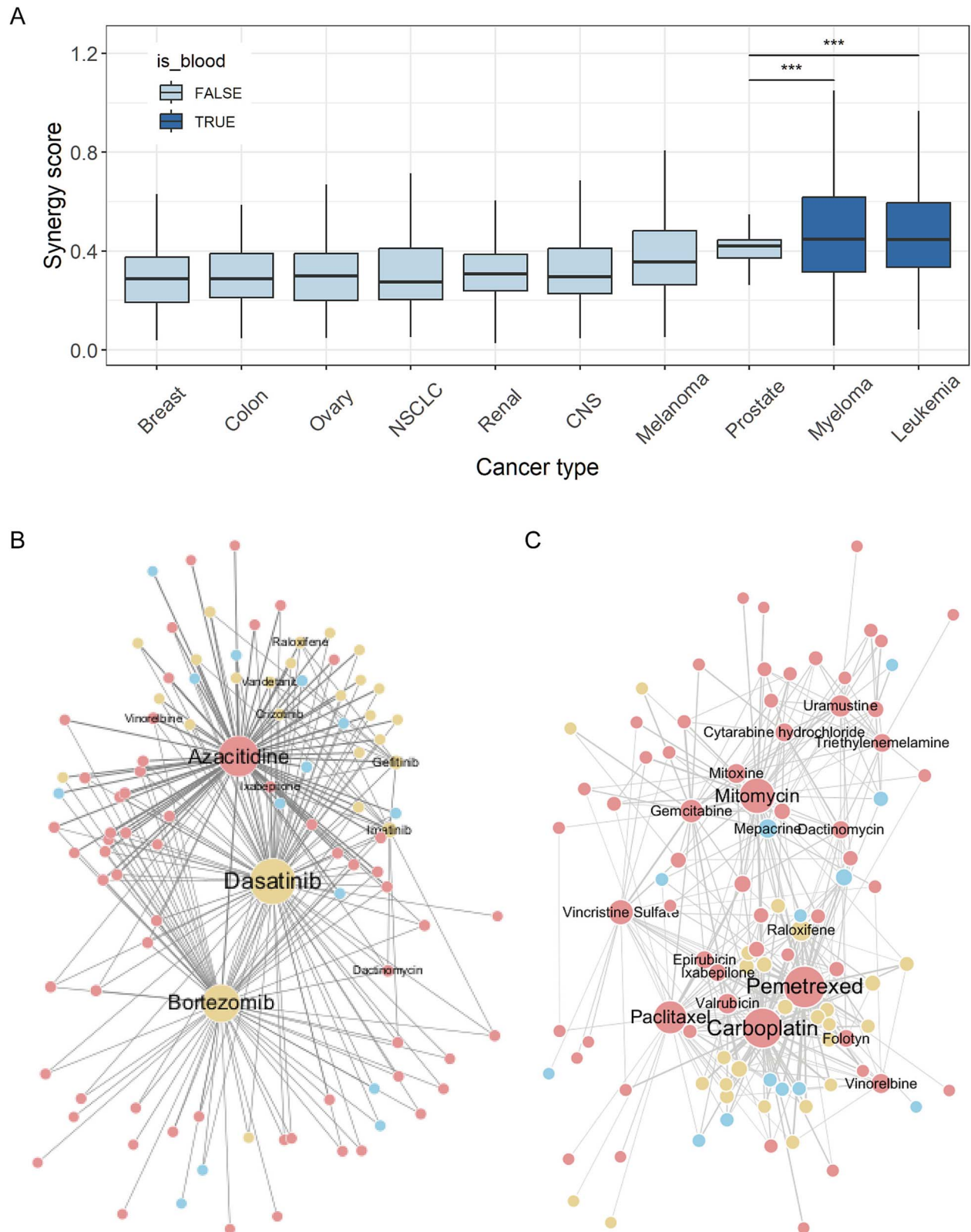
Figure 7. Cancer type–specific analyses. (A) The distribution of reachable synergy scores for each cancer type. Two navy boxes are for two blood cancer types, myeloma and leukemia, and the other boxes are for solid cancer types. (B, C) The drug synergism network whose nodes are drugs and the edges are the predicted synergism between drugs. The red node is a chemotherapy drug, the yellow node is a targeted therapy drug, and the blue node is a nonanticancer drug. The edge width is proportional to the reachable synergy score of the combination. (B) The leukemia cancer type. (C) The NSCLC cancer type.
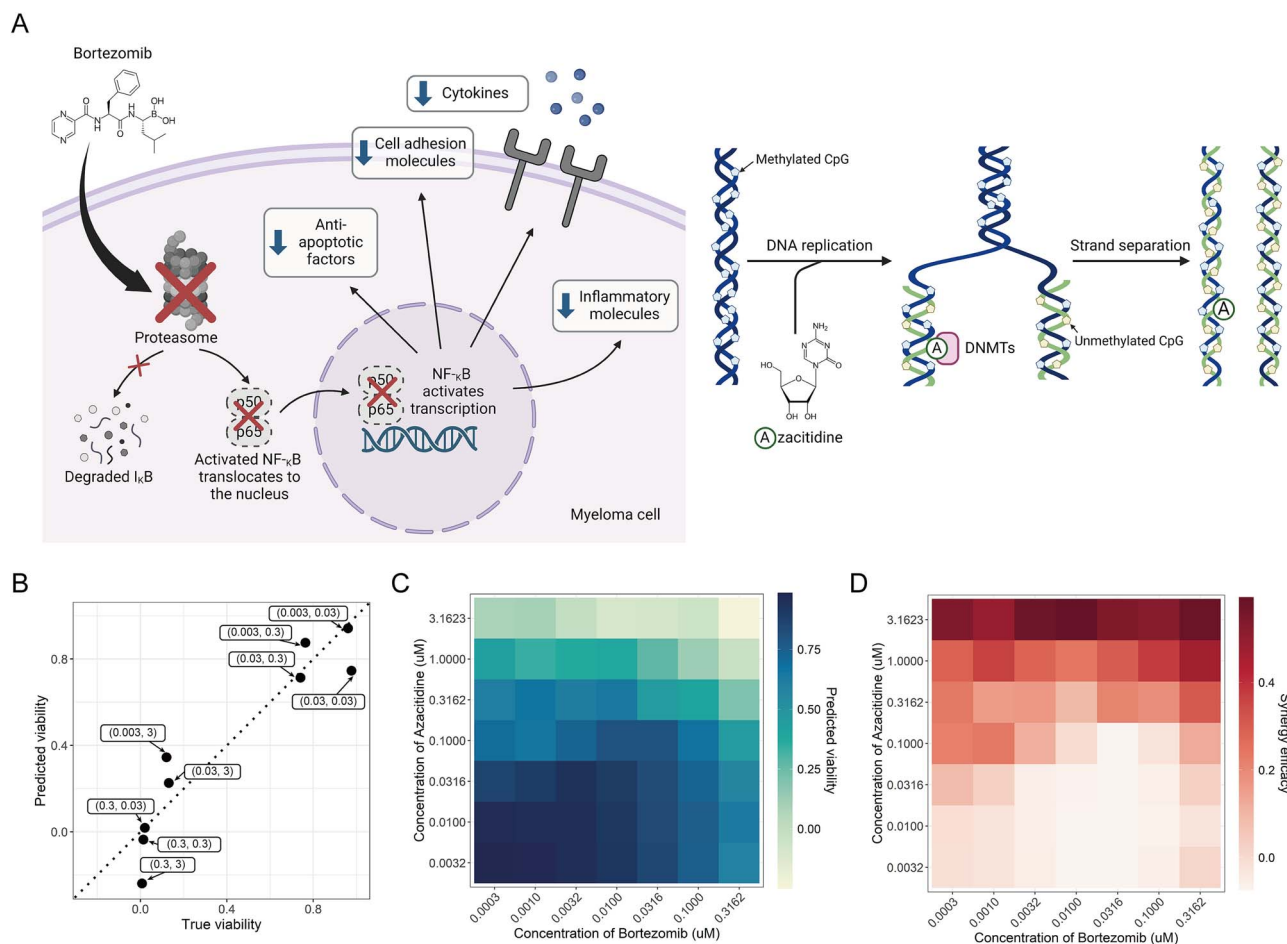
Figure 8. The predicted cell viability and components of RPMI-8226 + azacitidine + bortezomib decomposed by the DD-PRiSM. (A) The mechanism of action of two focused drugs, the proteasome inhibitor bortezomib and the DNA methyltransferase inhibitor azacitidine (created with BioRender. com). (B) The predicted result of RPMI-8226 + azacitidine + bortezomib with the Combination therapy model. (C) Predicted viability. (D) Synergy efficacy decomposed from the predicted viability.

is crucial for determining the synergism of drug combinations. For instance, a low IC50 does not necessarily translate into high drug efficacy as the maximum achievable effect may not be substantial enough to cure the patient. Similarly, synergy scores based on Loewe additivity or Bliss independence only indicate relative benefits, without providing a complete picture of the drug's true therapeutic potential. In contrast, cell viability data derived from the full dose–response curve offer a more direct and comprehensive view of whether an anticancer drug can achieve the desired therapeutic outcome. This concentration-dependent information is far more valuable for precision medicine as it reflects the drug's actual efficacy across a range of concentrations rather than being limited to a single, dose-independent metric like IC50 or synergy scores.

Notably, our combination therapy model outperformed other dose-dependent models such as IDACombo, comboFM, and comboLTR. In addition to its strong performance, DD-PRiSM offers the distinct advantage of being able to predict novel data points without the need for pre-existing monotherapy data or dependence on specific features defined during training, providing greater flexibility compared to previous models.

One of the most significant innovations of DD-PRiSM is its ability to decompose combination therapy efficacy into monotherapy efficacies and synergistic effects. This decomposition provides valuable insights into the mechanisms underlying drug synergies.

Our clustering analysis of synergy effects revealed that targeted therapy drugs tend to have synergistic effects with other drugs more frequently than chemotherapy or nonanticancer drugs. This finding could have important implications for the design of future combination therapies. Additionally, DD-PRiSM offers a guideline for personalized medicine by recommending combination therapies with optimal drug concentration pairs when a strong synergistic effect is present, while suggesting monotherapy if the synergy is either absent or not significant enough to justify potential side effects. This capability enhances the model's applicability in tailoring treatments to individual patients.

Our cancer type–related analysis yielded results consistent with previous discoveries, showing that the hematopoietic malignancies more readily exhibit synergistic effects compared to the solid tumors. This observation underscores the importance of considering cancer types when designing combination therapies. Furthermore, our drug synergism networks unveiled hub drugs for each cancer type that could serve as core agents for the combination therapy. Many of these predictions were supported by existing clinical trials, validating the potential of our approach to identify clinically relevant drug combinations.

Despite these strengths, it is important to acknowledge the limitations of DD-PRiSM. Like previous studies, our model showed relatively low performance on drugs that were not included in the training dataset. It is a well-known problem that phenotypic

response prediction models show performance drop on unexplored drugs, whether the model uses physicochemical descriptors [52–54] or graph-based features [8, 55, 56]. These phenomena imply the fact that we need more informative drug features for the phenotypic prediction.

In summary, we introduced DD-PRiSM, a drug combination response prediction model that integrates both the efficacy and mechanisms of individual monotherapies. The model can decompose the efficacy of a combination therapy into the contributions from each monotherapy and the synergistic interaction between them. DD-PRiSM showed its versatility across various cancer types and drug classes, highlighting distinct synergism patterns for each cancer type. We also reported a curated list of predicted synergistic combinations, which could be a valuable resource for future research and clinical studies. As we continue to refine and expand this approach, we anticipate that DD-PRiSM will become an invaluable resource for personalized medicine, enabling the identification of truly synergistic combinations tailored to individual patients and cancer types.

Key Points

- DD-PRiSM (Decomposition of Drug-Pair Response into Synergy and Monotherapy effect) uniquely integrates the monotherapy efficacy prediction of two drugs to forecast combination therapy outcomes, offering a more comprehensive approach than existing models.
- DD-PRiSM can decompose combination therapy efficacy into individual drug effects and synergistic interactions, providing unprecedented insights into the mechanisms of drug synergies.
- Our model outperforms current state-of-the-art dosage-dependent prediction methods, demonstrating remarkable accuracy in predicting combination therapy efficacies.
- Unlike previous models, DD-PRiSM can make predictions for unseen cell lines and drugs, significantly expanding its potential applications in combination therapy research.
- The ability of DD-PRiSM to predict synergistic combinations tailored to individual cancer types opens new avenues for personalized treatment strategies.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Acknowledgements

A biological illustration in Fig. 8A was created using BioRender.com.

Conflicts of interest: The authors declare no competing interests.

Funding

Data availability

All datasets used in this work were derived from sources in the public domain: NCI60, NCI-ALMANAC, O'Neil dataset, CCLE, and MSigDB. All source code from this work is available at https://github.com/GIST-CSBL/DD-PRiSM. The repository also includes links to download the raw datasets and the source code for preprocessing them.

Author contributions

References

1. Shyr ZA, Cheng Y-S, Lo DC. *et al.* Drug combination therapy for emerging viral diseases. *Drug Discov Today* 2021;**26**:2367–76. https://doi.org/10.1016/j.drudis.2021.05.008.
2. Jin H, Wang L, Bernards R. Rational combinations of targeted cancer therapies: Background, advances and challenges. *Nat Rev Drug Discov* 2023;**22**:213–34. https://doi.org/10.1038/s41573-022-00615-z.
3. Zhu S, Zhang T, Zheng L. *et al.* Combination strategies to maximize the benefits of cancer immunotherapy. *J Hematol Oncol* 2021;**14**:156. https://doi.org/10.1186/s13045-021-01164-5.
4. Mokhtari RB, Homayouni TS, Baluch N. *et al.* Combination therapy in combating cancer. *Oncotarget* 2017;**8**:38022. https://doi.org/10.18632/oncotarget.16723.
5. 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. https://doi.org/10.1158/0008-5472.CAN-17-0489.
6. Preuer K, Lewis RP, Hochreiter S. *et al.* DeepSynergy: Predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46. https://doi.org/10.1093/bioinformatics/btx806.
7. Kuru HI, Tastan O, Cicek AE. MatchMaker: A deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**:2334–44. https://doi.org/10.1109/TCBB.2021.3086702.
8. 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. https://doi.org/10.1093/bib/bbab390.
9. Wang T, Wang R, Wei L. AttenSyn: An attention-based deep graph neural network for anticancer synergistic drug combination prediction. *J Chem Inf Model* 2023;**64**:2854–62. https://doi.org/10.1021/acs.jcim.3c00709.
10. Loewe S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 1953;**3**:285–90.
11. Bliss CI. The toxicity of poisons applied jointly 1. *Ann Appl Biol* 1939;**26**:585–615. https://doi.org/10.1111/j.1744-7348.1939.tb06990.x.
12. Zagidullin B, Aldahdooh J, Zheng S. *et al.* DrugComb: An integrative cancer drug combination data portal. *Nucleic Acids Res* 2019;**47**:W43–51. https://doi.org/10.1093/nar/gkz337.
13. Ling A, Huang RS. Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action. *Nat Commun* 2020;**11**:5848. https://doi.org/10.1038/s41467-020-19563-6.

14. Julkunen H, Cichonska A, Gautam P. *et al.* Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nat Commun* 2020;**11**:6136. https://doi.org/10.1038/s41467-020-19950-z.

15. Wang T, Szedmak S, Wang H. *et al.* Modeling drug combination effects via latent tensor reconstruction. *Bioinformatics* 2021;**37**:i93–101. https://doi.org/10.1093/bioinformatics/btab308.

16. Palmer AC, Sorger PK. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell* 2017;**171**:1678–1691.e13. https://doi.org/10.1016/j.cell.2017.11.009.

17. Szedmak S, Cichonska A, Julkunen H. *et al.* A solution for large scale nonlinear regression with high rank and degree at constant memory complexity via latent tensor reconstruction, arXiv preprint arXiv:2005.01538 2020.

18. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;**6**:813–23. https://doi.org/10.1038/nrc1951.

19. O'Neil J, Benita Y, Feldman I. *et al.* An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016;**15**:1155–62. https://doi.org/10.1158/1535-7163.MCT-15-0843.

20. Landrum G. *RDKit: Open-Source Cheminformatics* 2010. https://www.rdkit.org/.

21. 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. https://doi.org/10.1038/nature11003.

22. Tsherniak A, Vazquez F, Montgomery PG. *et al.* Defining a cancer dependency map. *Cell* 2017;**170**:564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

23. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30. https://doi.org/10.1093/nar/28.1.27.

24. Subramanian A, Tamayo P, Mootha VK. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005;**102**:15545–50. https://doi.org/10.1073/pnas.0506580102.

25. Liberzon A, Birger C, Thorvaldsdóttir H. *et al.* The molecular signatures database hallmark gene set collection. *Cell Syst* 2015;**1**:417–25. https://doi.org/10.1016/j.cels.2015.12.004.

26. Vis DJ, Bombardelli L, Lightfoot H. *et al.* Multilevel models improve precision and speed of IC50 estimates. *Pharmacogenomics* 2016;**17**:691–700. https://doi.org/10.2217/pgs.16.15.

27. Jin I, Nam H. HiDRA: Hierarchical network for drug response prediction with attention. *J Chem Inf Model* 2021;**61**:3858–67. https://doi.org/10.1021/acs.jcim.1c00706.

28. Atmaja BT, Akagi M. Evaluation of error-and correlation-based loss functions for multitask learning dimensional speech emotion recognition. In: *J Phys Conf Ser*. Vol. **1896**. 2021, p. 012004. IOP Publishing.

29. Loshchilov I, Hutter F. Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 2017. https://doi.org/10.48550/arXiv.1711.05101.

30. Jiang L, Jiang C, Yu X. *et al.* DeepTTA: A transformer-based model for predicting cancer drug response. *Brief Bioinform* 2022;**23**:bbac100. https://doi.org/10.1093/bib/bbac100.

31. Virtanen P, Gommers R, Oliphant TE. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;**17**:261–72. https://doi.org/10.1038/s41592-019-0686-2.

32. Paszke A, Gross S, Massa F. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019;**32**. https://pytorch.org/ https://doi.org/10.1093/bib/bbac100.

33. Berend D, Xie X, Ma L. *et al.* Cats are not fish: Deep learning testing calls for out-of-distribution awareness. In: *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*. 2020, p. 1041–52. https://doi.org/10.1145/3324884.3416609.

34. Jaaks P, Coker EA, Vis DJ. *et al.* Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* 2022;**603**:166–73. https://doi.org/10.1038/s41586-022-04437-2.

35. Kong W, Midena G, Chen Y. *et al.* Systematic review of computational methods for drug combination prediction. *Comput Struct Biotechnol J* 2022;**20**:2807–14. https://doi.org/10.1016/j.csbj.2022.05.055.

36. Dhillon IS. Co-Clustering Documents and Words Using Bipartite Spectral Graph Partitioning. In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001. p. 269–274. https://doi.org/10.1145/502512.502550.

37. Pedregosa F, Varoquaux G, Gramfort A. *et al.* Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.

38. He L, Tang J, Andersson EI. *et al.* Patient-customized drug combination prediction and testing for T-cell prolymphocytic leukemia patients. *Cancer Res* 2018;**78**:2407–18. https://doi.org/10.1158/0008-5472.CAN-17-3644.

39. Pomeroy AE, Schmidt EV, Sorger PK. *et al.* Drug independence and the curability of cancer by combination chemotherapy. *Trends Cancer* 2022;**8**:915–29. https://doi.org/10.1016/j.trecan.2022.06.009.

40. Slayton WB, Schultz KR, Kairalla JA. *et al.* Dasatinib plus intensive chemotherapy in children, adolescents, and young adults with Philadelphia chromosome–positive acute lymphoblastic leukemia: Results of children's oncology group trial AALL0622. *J Clin Oncol* 2018;**36**:2306–14. https://doi.org/10.1200/JCO.2017.76.7228.

41. Marcucci G, Geyer S, Laumann K. *et al.* Combination of dasatinib with chemotherapy in previously untreated core binding factor acute myeloid leukemia: CALGB 10801. *Blood Adv* 2020;**4**:696–705. https://doi.org/10.1182/bloodadvances.2019000492.

42. DiNardo CD, Daver N, Jabbour E. *et al.* Sequential azacitidine and lenalidomide in patients with high-risk myelodysplastic syndromes and acute myeloid leukaemia: A single-arm, phase 1/2 study. *Lancet Haematol* 2015;**2**:e12–20. https://doi.org/10.1016/S2352-3026(14)00026-X.

43. Montalban-Bravo G, Huang X, Jabbour E. *et al.* A clinical trial for patients with acute myeloid leukemia or myelodysplastic syndromes not eligible for standard clinical trials. *Leukemia* 2017;**31**:318–24. https://doi.org/10.1038/leu.2016.303.

44. Horton TM, Gannavarapu A, Blaney SM. *et al.* Bortezomib interactions with chemotherapy agents in acute leukemia in vitro. *Cancer Chemother Pharmacol* 2006;**58**:13–23. https://doi.org/10.1007/s00280-005-0135-z.

45. Stathopoulos GP, Dimitroulis J, Toubis M. *et al.* Pemetrexed combined with paclitaxel in patients with advanced or metastatic non-small-cell lung cancer: A phase I-II trial. *Lung Cancer* 2007;**57**:66–71. https://doi.org/10.1016/j.lungcan.2007.02.003.

46. Gubens MA, Chuang JC, Akerley W. *et al.* A pooled analysis of advanced nonsquamous non-small cell lung cancer patients with stable treated brain metastases in two phase II trials receiving bevacizumab and pemetrexed as second-line therapy. *J Thorac Dis* 2018;**10**:219. https://doi.org/10.21037/jtd.2017.12.30.

47. Kim YH, Hirabayashi M, Togashi Y. *et al.* Phase II study of carboplatin and pemetrexed in advanced non-squamous,

non-small-cell lung cancer: Kyoto thoracic oncology research group trial 0902. *Cancer Chemother Pharmacol* 2012;**70**:271–6. https://doi.org/10.1007/s00280-012-1910-2.

48. Schuette WH, Gröschel A, Sebastian M. *et al.* A randomized phase II study of pemetrexed in combination with cisplatin or carboplatin as first-line therapy for patients with locally advanced or metastatic non–small-cell lung cancer. *Clin Lung Cancer* 2013;**14**: 215–23. https://doi.org/10.1016/j.cllc.2012.10.001.

49. Gervais R, Robinet G, Clément-Duchêne C. *et al.* Pemetrexed and carboplatin, an active option in first-line treatment of elderly patients with advanced non-small cell lung cancer (NSCLC): A phase II trial. *Lung Cancer* 2013;**80**:185–90. https://doi.org/10.1016/j.lungcan.2013.01.008.

50. Kiziltepe T, Hideshima T, Catley L. *et al.* 5-azacytidine, a DNA methyltransferase inhibitor, induces ATR-mediated DNA double-strand break responses, apoptosis, and synergistic cytotoxicity with doxorubicin and bortezomib against multiple myeloma cells. *Mol Cancer Ther* 2007;**6**:1718–27. https://doi.org/10.1158/1535-7163.MCT-07-0010.

51. Hu X, Xuan H, Du H. *et al.* Down-regulation of CD9 by methylation decreased bortezomib sensitivity in multiple myeloma. *PloS One* 2014;**9**:e95765. https://doi.org/10.1371/journal.pone.0095765.

52. Preto AJ, Matos-Filipe P, Mourão J. *et al.* SYNPRED: Prediction of drug combination effects in cancer using different synergy metrics and ensemble learning. *GigaScience* 2022;**11**:giac087. https://doi.org/10.1093/gigascience/giac087.

53. Li Y, Hostallero DE, Emad A. Interpretable deep learning architectures for improving drug response prediction performance: Myth or reality? *Bioinformatics* 2023;**39**:btad390. https://doi.org/10.1093/bioinformatics/btad390.

54. Tang Y-C, Gottlieb A. Explainable drug sensitivity prediction through cancer pathway enrichment. *Sci Rep* 2021;**11**:3128. https://doi.org/10.1038/s41598-021-82612-7.

55. Cheng X, Dai C, Wen Y. *et al.* NeRD: A multichannel neural network to predict cellular response of drugs by integrating multidimensional data. *BMC Med* 2022;**20**:368. https://doi.org/10.1186/s12916-022-02549-0.

56. 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. https://doi.org/10.1093/bib/bbac302.