Zhao *et al. Journal of Cheminformatics* (2024) 16:41
https://doi.org/10.1186/s13321-024-00839-8

Journal of Cheminformatics

RESEARCH Open Access

PermuteDDS: a permutable feature fusion network for drug-drug synergy prediction

Xinwei Zhao[1], Junqing Xu[2], Youyuan Shui[1], Mengdie Xu[1], Jie Hu[1,3], Xiaoyan Liu[5], Kai Che[6,7], Junjie Wang[1,3*] and Yun Liu[1,3,4*]

Abstract

Motivation Drug combination therapies have shown promise in clinical cancer treatments. However, it is hard to experimentally identify all drug combinations for synergistic interaction even with high-throughput screening due to the vast space of potential combinations. Although a number of computational methods for drug synergy prediction have proven successful in narrowing down this space, fusing drug pairs and cell line features effectively still lacks study, hindering current algorithms from understanding the complex interaction between drugs and cell lines.

Results In this paper, we proposed a Permutable feature fusion network for Drug-Drug Synergy prediction, named PermuteDDS. PermuteDDS takes multiple representations of drugs and cell lines as input and employs a permutable fusion mechanism to combine drug and cell line features. In experiments, PermuteDDS exhibits state-of-the-art performance on two benchmark data sets. Additionally, the results on independent test set grouped by different tissues reveal that PermuteDDS has good generalization performance. We believed that PermuteDDS is an effective and valuable tool for identifying synergistic drug combinations. It is publicly available at https://github.com/littleweilazy/PermuteDDS.

Scientific contribution First, this paper proposes a permutable feature fusion network for predicting drug synergy termed PermuteDDS, which extract diverse information from multiple drug representations and cell line representations. Second, the permutable fusion mechanism combine the drug and cell line features by integrating information of different channels, enabling the utilization of complex relationships between drugs and cell lines. Third, comparative and ablation experiments provide evidence of the efficacy of PermuteDDS in predicting drug-drug synergy.

Keywords Drug combination, Permutable feature fusion, Synergistic effect, Cancer

*Correspondence:
Junjie Wang
junjie2021@njmu.edu.cn
Yun Liu
liuyun@njmu.edu.cn
[1] Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China
[2] The Second Clinical Medical School, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China
[3] Institute of Medical Informatics and Management, Nanjing Medical University, 101 Longmian Avenue, Nanjing 210029, Jiangsu, China
[4] Department of Information, the First Affiliated Hospital, Nanjing Medical University, No. 300 Guang Zhou Road, Nanjing 210029, Jiangsu, China
[5] Faculty of Computing, Harbin Institute of Technology, No. 92 West Da Zhi St, Harbin 150001, Heilongjiang, China
[6] Xi'an Aeronautics Computing Technique Research Institute, AVIC, No. 156, TaiBai Nroth Road, Xi'an 710068, Shanxi, China
[7] Aviation Key Laboratory of Science and Technology on Airborne and Missleborne Computer, Xi'an 710065, Shanxi, China

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Single-agent targeted therapies have been widely utilized in clinical cancer treatments. However, the clinical efficacy of monotherapy remains limited due to the biological complexity of tumors and the presence of pre-existing or acquired drug resistance mechanisms [1]. In contrast, combination therapies involving the simultaneous administration of multiple drugs have shown promising results in cancer treatment [2]. Compared to monotherapy, combination therapies have the potential to enhance treatment efficacy, reduce the dose-limiting toxicity associated with single agents, and overcome drug resistance [3]. Despite the potential benefits, it is important to note that not all drug combinations exhibit synergistic effects, and some combinations may even result in antagonistic interactions [4]. Therefore, it is essential to accurately determine the interaction between drug combinations to specific diseases.

Traditionally, the discovery of drug combinations heavily relied on clinical trials, posing challenges in terms of time consumption and cost-effectiveness due to the extensive number of potential drug combinations [5]. An alternative method is high-throughput screening (HTS), which enables automated testing of chemical and biological compounds against specific biological targets [6]. HTS techniques have significantly reduced the time required for identifying potential drug combinations. However, it is important to note that HTS has limitations in revealing the in vivo action modes of drug molecules [7]. Moreover, the substantial increase in the number of available drugs has rendered it impractical to comprehensively test the entire combinatorial space using HTS [8]. The sheer magnitude of potential drug combinations poses a challenge in terms of feasibility and resource utilization.

Computational methods, particularly machine learning models, have emerged as powerful tools for exploring the vast synergistic space of drug combinations [9]. These machine learning-based models can be broadly categorized into two types: classical machine learning and deep learning. Classical machine learning models, such as random forest [10], support vector machine (SVM) [11], extreme gradient boosting (XGBoost) [12], have been widely utilized in drug synergy prediction. These models leverage genomic information from cancer cells, physiochemical properties of drugs, and drug-cell interaction data to predict drug synergy [13]. For instance, Jeon et al. [14] employed random forest and SVM algorithms to predict the synergistic effects of anticancer drug combinations by incorporating monotherapy response and synthetic lethality information. However, this approach heavily relies on the given dataset, and the model may struggle to accurately predict the features of drug combinations without prior knowledge. Another study by He et al. [15] proposed an XGBoost-based model that predicts the selective synergistic effects of cancer by utilizing distinct single compound sensitivity curves between patient cells and healthy controls. This approach aimed to minimize potential toxicity associated with drug combinations.

Recently, deep learning has gained increasing popularity in drug development and discovery. Compare to classical machine learning methods, deep learning algorithms have better generalization performance, making it efficient in handling large drug combination datasets. Deep learning-based methods typically frame the challenge of drug synergy prediction as a regression task, aiming to predict the quantitative synergy scores. These methods can be categorized into three groups: fingerprint-based methods, SMILE-based methods and graph-based methods. Fingerprint-based methods take molecular fingerprint (also called descriptor) as input. Preuer et al. [9] proposed DeepSynergy, a feedforward neural network-based deep learning model for drug synergy score prediction that employs molecular fingerprints and gene expression as inputs. The performance of DeepSynergy demonstrated significant improvement over classical machine learning models. Kuru et al. [16] developed a MatchMaker model for predicting synergistic drug combinations using chemical descriptors generated by ChemoPy [17] and the expression profiles of landmark genes as input. Lin et al. [18] amalgamated molecular fingerprint information with drug induced gene expression profiles to capture drug cell responses to reveal the mechanisms of biological synergistic effects, and using deep forests as feature learning models. Hosseini et al. [19] proposed CCSynergy, used a feed forward network to obtain the fusion feature of drug fingerprints and mutiple cell line representations for durg syergy prediction. SMILE-based methods select simplified molecular-input line-entry system (SMILES) [20] as drug representations. Kim et al. [21] utilized a muti-head attention mechanism and convolutional neural networks (CNN) to encode drug SMILES. Graph-based methods that extract features from molecular graphs have the potential to capture structural information about the molecules [22]. Wang et al. [23] proposed DeepDDS, a model that utilizes a graph convolutional network (GCN) and attention mechanism to compute drug embedding vectors, which are integrated with cell line gene expression data to predict drug synergy. Liu et al. [24] proposed HypergraphSynergy, a method employing hypergraph representation learning to predict anti-cancer drug synergy by considering drugs and cell lines as nodes and representing drug pair-cell line triplets as hyperedges.

The previously mentioned deep learning-based methods have exhibited impressive performance in predicting

drug synergy, yet there is still room for further improvement. One limitation is that these methods typically only use single type drug representation, thereby failing to provide a comprehensive description of drugs. Moreover, most of these methods simply fuse drug and cell line features through simple operations such as concatenation or addition, neglecting dynamic synergies between drug pairs and cell lines.

In light of the limitations associated with existing approaches for predicting drug synergy, we proposed a Permutable feature fusion network for Drug-Drug Synergy prediction, namely PermuteDDS. Our method addresses the challenge of limited representation types by considering multiple representations of drugs and cell lines. We employed the one-dimensional CNN to learn and extract high-latent features from these representations. Then, we designed a permutable fusion mechanism to combine the drug and cell line features, allowing us to capture the complex relationships between drug combinations and cell lines. Finally, we conducted experiments on two benchmark datasets, and the results compared to state-of-the-art methods indicated that PermuteDDS is an effective model for drug synergy prediction. In summary, the contributions of our work are as follows:

- We propose PermuteDDS, a deep learning model for drug-drug synergy prediction with multiple input representations.
- Our major contribution lies in the application of the permutable fusion mechanism to represent the interactions between drug combinations and cell lines.
- We conduct comprehensive experiments to demonstrate the effectiveness of our proposed model. We present ablations and analysis to gain a deeper understanding of the model's behavior.

Materials and methods

Data description

We collected Drug-Drug Synergy (DDS) data, molecular structures of drugs, the expression profiles and mutation information of cell lines from various public databases as below.

- **Drug-Drug Synergy datasets.** The Drug-Drug Synergy (DDS) data were obtained from two released large-scale oncology screening datasets, namely O'Neil [25] and NCI-ALMANAC [26]. The O'Neil dataset contains 22 737 samples estimated by Loewe synergy score [27] consisting of two chemicals and a cell line, covering 38 unique drugs and 39 diverse caner cell lines. The NCI-ALMANAC dataset comprises 304 549 samples evaluated by the ComboScores of pairwise combinations of 104 FDA-approved anticancer drugs against a panel of 60 well-characterized human tumor cell lines.
- **Drug synergy scores.** The Loewe synergy score incorporates the concepts of sham combination and dose equivalence. Let A and B denote two drugs, and x_A and x_B represent their respective doses. The drug synergy score can then be computed as follows [27]:

$$Y_{Loewe} = \frac{R_{\min} + R_{\max}\left(\frac{x_A + x_B}{m}\right)^{\lambda}}{1 + \left(\frac{x_A + x_B}{m}\right)^{\lambda}} \tag{1}$$

where Y_{Loewe} is a continuous value, $R_{\min}$ and $R_{\max}$ represents the maximum and minimum drug response, respectively. λ is the shape parameter and m is the dose that produce midpoint between $R_{\min}$ and $R_{\max}$. The determination of combination benefit, denoted as 'ComboScore' [26], relies on a modification of Bliss independence [28]. Let $Y_i^{A_pB_q}$ represent the growth fraction for the i-th cell line exposed to the p-th concentration of drug A and the q-th concentration of drug B. Similarly, let $Z_i^{A_pB_q}$ denote the expected growth fraction for the combination. The final continuous combination score $Y_{ComboScore}$ for the cell line and the drug combination can then be computed as the sum of the differences between expected and observed growth fractions:

$$Y_{ComboScore} = \sum_{p,q} Y_i^{A_pB_q} - Z_i^{A_pB_q} \tag{2}$$

- **Molecular structures of drugs.** The SMILES of the drugs were obtained from PubChem database [29], based on which the chemical and structure information of a drug can be converted to fingerprints with RDKit [30] and MAP4 project.[1]
- **Gene expression and gene mutation of cell lines.** The gene expression and gene mutation of cell lines were obtained from Genomics of Drug Sensitivity in Cancer (GDSC) database,[2]. where 1000 human cancer cell lines were characterised and screened [31]. To enhance the representation of cell lines, we identified significant genes by referencing The Library of Integrated Network-Based Cellular Signatures (LINCS) project [32]. The LINCS project offers a meticulously curated set of approximately 1000 genes, known as the 'landmark gene set', which captures 80% of the information from Connectivity Map data [33]. We selected genes that intersected between GDSC gene expression profiles and the landmark set,

[1] https://github.com/reymond-group/MAP4/tree/master.

[2] https://www.cancerrxgene.org.

as well as gene mutation profiles. Finally, 899 genes were chosen for expression profiles, and 968 genes were selected for mutation profiles.

We preprocessed the two dataset by eliminating the drugs lacking SMILES and the cell lines devoid of gene expressions or gene mutations. The resulting refined O'Neil dataset comprised 18 950 samples with Loewe synergy scores across 38 drugs and 39 cell lines, while the processed NCI-AlMANAC dataset consisted of 74 139 instances with NCI ComboScores (a Bliss independence modification) across 87 drugs and 55 cell lines.

Molecular fingerprints

Molecular fingerprints (also referred to as descriptors) is a prevalent method of drug representation. Within this approach, each molecule is encoded as a binary string that delineate the presence or absence of specific substructure fragments or properties within the molecule. A value of 0 corresponds to absence of the particular feature, while a value of 1 denotes its existence. Moreover, the fingerprints exhibit uniqueness; specifically, one molecule possesses one sequence for a particular fingerprints type, yet it can have numerous diverse fingerprints types [22]. The three different kinds of fingerprints employed in this article are described below.

- **Hashed topological-torsion (HashTT)** is a kind of topological fingerprints [34]. The fragments employed in HashTT fingerprints are linear sequence comprising four consecutively bonded non-hydrogen atoms. The label for each fragment is determined based on its atomic type, the number of non-hydrogen branches attached to it, and its number of 7r electrons of each atom. These labels then undergo a hashing process to generate the HashTT fingerprints.
- **MinHashed atom-pair fingerprint up to a diameter of four bonds (MAP4)** [35] is a novel fingerprint that integrates both substructure fingerprints and atom-pair fingerprints. Given a molecule SMILE sentence, the circular substructures with radii of r = 1 and r = 2 bonds around individual atoms within an atom-pair configuration are encoded as two pairs of SMILES. Each pair is subsequently combined with the topological distance separating the central atoms. These atom-pair molecular shingles undergo a hashing process, and the resulting set of hashes is subjected to the MinHashed technique, which lead to the creation of the MAP4 fingerprint.
- **Molecular ACCess System (MACCS)** [36] keys are one of the most commonly used structural fingerprints. In this fingerprint, one molecule is encoded as a 166-bit binary string, where each bit corresponds to a predefined structural fragment (e.g. 3-element ring).

Overview of PermuteDDS

PermuteDDS is an end-to-end framework for predicting drug combinations, presented in Fig. 1. Specifically, Fig. 1(A) illustrates the pipeline implemented by PermuteDDS to derive the synergy scores. Our framework contains three subnetworks called Fingerprint Specific Networks (FSNs), all of which share the same architecture shown in Fig. 1(B). Each FSN takes a drug pair of a certain fingerprint and a cell line pair including its gene expression and gene mutation profiles as inputs. Then, the CNNs are used to extract drug and cell line features, and the Permute-MLP module (Fig. 1(C)) is proposed to obtain the fusion synergy feature, which is the output of FSN. The three FSNs output three distinct synergy features, each of which is propagated through a specific prediction module to generate a synergy score. The final outputs of PermuteDDS is the average of the synergy scores predicted by these prediction modules.

Fingerprint specific subnetwork (FSN)

Given a drug fingerprint pair and a cell line pair as inputs, the FSN encompasses three essential processes, including drug features extraction, cell line features extraction and feature fusion, which integrates the extracted drug and cell line features. The output of FSN is a synergy feature. A comprehensive overview of these processes is provided below.

Drug features To extract the drug fingerprint features, we used the one-dimensional convolutional neural network (Conv1D), where 1D array-like kernel convolves along a single dimension and identify the patterns from fingerprint information. Here, we considered three different fingerprints include Hashtt(t), MAP4(p) and MACCS(s). We conducted a three-layer Conv1D model, as illustrated in Fig. 2. An input fingerprint sequence is represented as $\mathbf{X}_d \in \mathbb{R}^{L\times 1}$ in which L denotes the length, 1 denotes the dimension and $d \in \{t, p, s\}$. Let $\mathbf{X}_{i:i+j}$ refer to the concatenation of tokens $\mathbf{X}_i, \mathbf{X}_{i+1}, ..., \mathbf{X}_{i+j}$ in the sequence, a feature $z_i \in \mathbb{R}^k$ can be generated from the window of tokens $\mathbf{X}_{i:i+h-1}$ by

$$z_i = \mathrm{GELU}\left(\mathbf{W}_\mathrm{k} \cdot \mathbf{X}_\mathrm{i:i+h-1} + \mathrm{b}_\mathrm{k}\right), \tag{3}$$

where GELU [37] is an activation function, $\mathbf{W}_k \in \mathbb{R}^{h\times 1\times k}$ is a k-channel (here is 8) filter applied to a window of h token, $b_k \in \mathbb{R}$ is a bias term and $z_i \in \mathbb{R}^k$ is a feature generated from the window of tokens $\mathbf{X}_{i:i+h-1}$. Then, the filter is applied to each possible window of tokens to produce a feature map

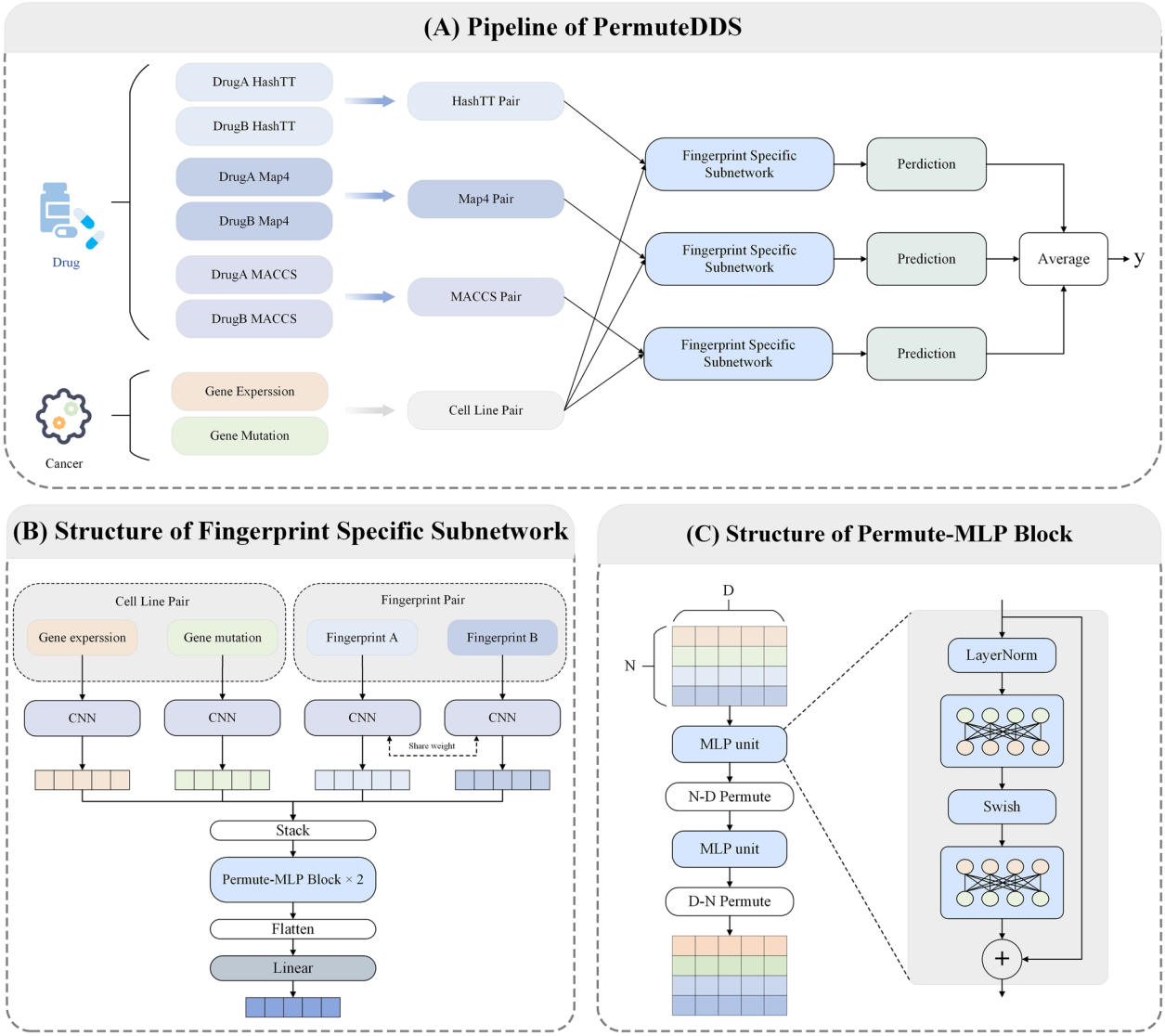

Fig. 1 The overview of the PermuteDDS framework

$$\mathbf{z} = \mathrm{Concat}(z_1, z_2, ..., z_L), \tag{4}$$

where *Concat* refers to the concatenation process and $\mathbf{z} \in \mathbb{R}^{L \times k}$. Subsequently, we conducted similar operations as above with different filter channels (16 and 32 respectively) in the next layers and then got an output $\mathbf{z}' \in \mathbb{R}^{L \times k'}$ in which k' denote the output channel (here is 32) of the third layer. To obtained the final drug features with expected dimension D, we conducted a projection process as follow:

$$\mathbf{Z}_d = \mathbf{W}_d \cdot \mathrm{Flatten}(\mathbf{z}') + \mathrm{b_d}, \tag{5}$$

where F*latten* refers to the flattening process, $\mathbf{W}_d \in \mathbb{R}^{Lk' \times D}$ is the weight matrix, b_d is the bias term and $\mathbf{Z}_d \in \mathbb{R}^D$ is the output drug features. Then, given a drug pair (i, j), the pairwise features can be represented as $(\mathbf{Z}_{di}, \mathbf{Z}_{dj})$, where $\mathbf{Z}_{di} \in \mathbb{R}^D$ and $\mathbf{Z}_{dj} \in \mathbb{R}^D$.

Cell line features To derive cell line features, we employed two distinct three-layer CNNs individually for gene expression (e) and gene mutation (m) data. Consistent with the fingerprint sequence, an initial gene expression sequence can be represented as $\mathbf{X}_e \in \mathbb{R}^{l \times 1}$ in which l denotes the number of selected genes and 1 is the dimension. We then conducted the same operations described in the Eq. 3, 4 and 5 to obtained the gene expression features $\mathbf{Z}_e \in \mathbb{R}^D$ with D represents the output dimension, which aligns with the dimension of

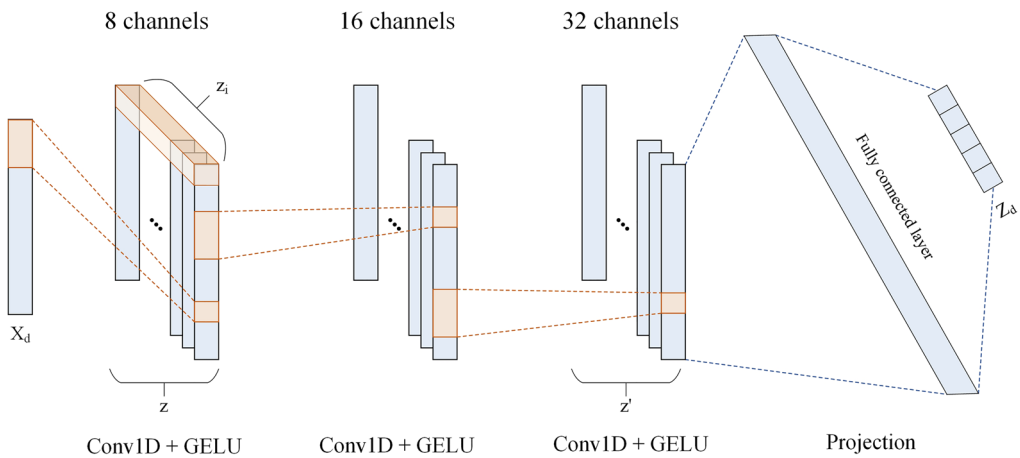

Fig. 2 The architecture of the three-layer Conv1D

drug features. Similarly, the extracted gene mutation features are represented as $\mathbf{Z}_m \in \mathbb{R}^D$.

Feature fusion Recent works [38, 39] have demonstrated the superior performance of Multilayer Perceptron (MLP) in feature fusion tasks. Motivated by these findings, we conducted a MLP-based method to extract fusion feature of drug fingerprints and cell lines. Given a triplet of drug pairs and cell line, a feature $\mathbf{H}_d \in \mathbb{R}^{N\times D}$ is constructed by

$$\mathbf{H}_d = \mathrm{Stack}(\mathbf{Z}_{\mathrm{di}}, \mathbf{Z}_{\mathrm{dj}}, \mathbf{Z}_{\mathrm{e}}, \mathbf{Z}_{\mathrm{m}}). \tag{6}$$

Here, N corresponds to the value 4, reflecting the number of features utilized in the stacking process. We then employed two Permute-MLP blocks taken the feature $\mathbf{H_d}$ as input. A diagrammatic illustration of the Permute-MLP block is presented in Fig. 1(C). First, we conducted a MLP unit along the D-channel, which can be formulated as follows:

$$\widehat{\mathbf{H}}_d = \mathbf{H}_d + \mathrm{LN}(\mathrm{Swish}(\mathbf{H}_{\mathrm{d}}\mathbf{W}^1_{\mathrm{dD}})\mathbf{W}^2_{\mathrm{dD}}), \tag{7}$$

where Swish [40] is an activation function, LN refers to LayerNorm [41], $\mathbf{W}^1_{dD}, \mathbf{W}^2_{dD} \in \mathbb{R}^{D\times D}$ are the weight matrices and $\widehat{\mathbf{H}}_d \in \mathbb{R}^{N\times D}$. Then, we perform a N-D permutation operation with respect to $\widehat{\mathbf{H}}_d$, yielding $\widehat{\mathbf{H}}^\top_d \in \mathbb{R}^{D\times N}$, which is serve as the input to the next MLP unit along the N-channel. This process can be described as

$$\overline{\mathbf{H}}_d = \widehat{\mathbf{H}}^\top_d + \mathrm{LN}(\mathrm{Swish}(\widehat{\mathbf{H}}^\top_{\mathrm{d}}\mathbf{W}^1_{\mathrm{dN}})\mathbf{W}^2_{\mathrm{dN}}), \tag{8}$$

where $\overline{\mathbf{H}}_d \in \mathbb{R}^{D\times N}$, $\mathbf{W}^1_{dN} \in \mathbb{R}^{N\times D}$ and $\mathbf{W}^2_{dN} \in \mathbb{R}^{D\times N}$. To recover the original dimensional information to $\mathbf{H_d}$, we only need to perform a D-N permutation operation on $\overline{\mathbf{H}}_d$, outputting $\overline{\mathbf{H}}^\top_d \in \mathbb{R}^{N\times D}$, which is the input to the next Permute-MLP block. Similarly, we conducted the same operations as above in the second block and obtained a fusion feature $\widetilde{\mathbf{H}}_d \in \mathbb{R}^{N\times D}$. Furthermore, we conducted projection process to obtain the final synergy feature

$$\mathbf{M}_d = \mathrm{GELU}(\mathrm{Flatten}(\widetilde{\mathbf{H}}_{\mathrm{d}})\mathbf{W}^1_{\mathrm{dproj}})\mathbf{W}^2_{\mathrm{dproj}}, \tag{9}$$

where $\mathbf{W}^1_{dproj} \in \mathbb{R}^{ND\times\widehat{D}}$ with hidden size $\widehat{D}$, $\mathbf{W}^2_{dproj} \in \mathbb{R}^{\widehat{D}\times\widetilde{D}}$ with the output dimension $\widetilde{D}$, and $\mathbf{M}_d \in \mathbb{R}^{\widetilde{D}}$ is the output of $\mathrm{FSN_d}$.

Predicting the synergistic effect

The objective of drug synergy prediction is to derive a synergy score for a given drug pair and cell line trio. The process outlined above results in the generation of three distinct synergy features, denoted as $\mathbf{M}_t$, $\mathbf{M}_p$ and $\mathbf{M}_s$, through the utilization of the respective modules $\mathrm{FSN_t}$, $\mathrm{FSN_p}$, and $\mathrm{FSN_s}$. Each of the generated features was then propagated to a specific prediction module, which consisted of two linear transformations with a GELU activation in between, and output a synergy score $\hat{y}_d \in \mathbb{R}$. In other words, we got three different synergy scores denoted as $\hat{y}_t$, $\hat{y}_p$ and $\hat{y}_s$. Given a drug-drug-cell line trio, the final predicted synergy score $\hat{y}$ can be calculated as

$$\hat{y} = \frac{1}{3}(\hat{y}_t + \hat{y}_p + \hat{y}_s) \tag{10}$$

Then, the mean square error (MSE) is adopted as the loss function to train the model, which is defined as:

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{11}$$

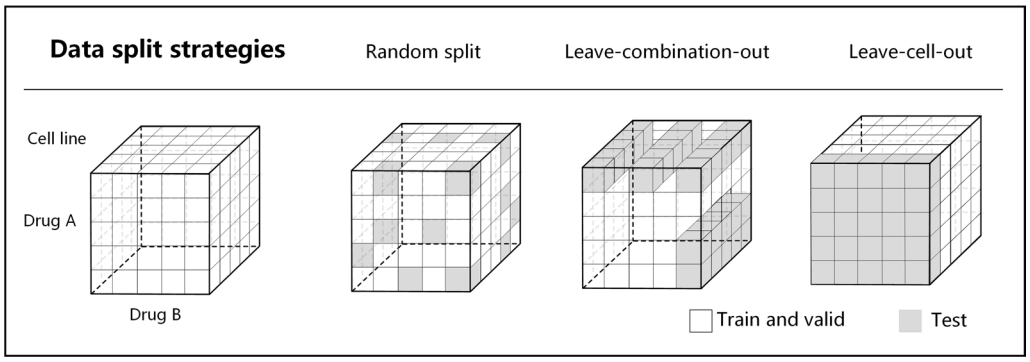

Fig. 3 Three different data split strategies. The white parts are used as training and validation data. The grey parts indicate testing data

where n represents the number of training set and y_i is the real score of a given trio.

Experimental setup

Data split strategies

We first randomly selected 90% samples from each dataset to conduct a 5-fold cross-validation strategy. To benchmark the performance of graph-fp under different situations, we considered three different strategies shown in Fig. 3:

- **Random-split**. The training samples were randomly partitioned into five equal folds. Four of these folds were utilized as the training set, while the remaining one was designated as the test set.
- **Leave-cell-out**. The cell line set was randomly partitioned into five equal folds. Samples containing the cell lines in four of these folds were used as the training set, while the remaining samples were assigned to the test set. This ensured that the test set contained only cell lines that were not included in the training set.
- **Leave-combination-out**. The set of drug combinations was randomly partitioned into five equal folds. Samples containing the drug combinations in four of these folds were used as the training set, while the remaining samples were used as the test set, ensuring that the test set only contained unseen drug combinations that were not present in the training set.

Furthermore, to verify the generalization ability of our method, we used the remaining 10% samples as the independent test set. The models were first trained and evaluated on the cross validation set, and then tested on the independent test set.

Baselines

To evaluate the performance of our model, we compared it with several competitive deep learning methods.

- **HypergraphSynergy** [24] employed Hypergraph Neural Networks (HGNN) to predict drug synergy with hypergraphs as input. In this method, drugs and cell lines are represented as nodes, while synergistic drug-drug-cell line triplets are represented as hyperedges. We reproduced the results of HypergraphSynergy and obtained the remaining results as reported in the HypergraphSynergy paper.
- **DeepSynergy** [9] takes molecular chemistry and cell line gene expression as input and used a feed forward neural network to predict synergy scores.
- **DTF** [42] utilized a tensor factorization to decompose drug synergy matrix and the results of the tensor decomposition are used as features to train the DNN model for drug synergy prediction.
- **CombFM** [43] used a higher-order factorization machine to predict synergy scores with a higher-order tensor as input which is compiled by drugs, drug concentrations and cell lines.
- **Celebi's method** [44] integrate drug synergy features with multiple biological and chemical properties' features and employed an XG-Boost model to predict synergy scores.

Evaluation metrics

We regarded the drug synergy prediction as a regression task, which objective is to predict the quantitative synergy scores of drug combinations. The regression results were evaluated by three metrics: the root man squared error (RMSE), the coefficient of determination (R^2) and Pearson's Correlation Coefficient (PCC). These evaluation metrics are calculated as follows:

Table 1 Performance comparison on the O'Neil dataset. Bold values indicate the best performance

	Random split			Leave-cell-out			Leave-combination-out		
	RMSE	R^2	PCC	RMSE	R^2	PCC	RMSE	R^2	PCC
PermuteDDS	**13.721**	**0.641**	**0.801**	19.668	0.243	0.522	**16.152**	**0.501**	**0.709**
HypergraphSynergy	14.727	0.586	0.775	**19.537**	**0.252**	0.533	17.346	0.420	0.656
DeepSynergy	14.87	0.584	0.765	23.890	0.195	0.426	17.28	0.433	0.663
ComboFM	16.86	0.451	0.702	20.820	0.142	0.396	18.62	0.376	0.635
DTF	14.73	0.594	0.775	21.110	0.132	**0.535**	17.37	0.429	0.671
Celebi's method	16.34	0.5	0.708	20.6	0.179	0.473	19.1	0.309	0.572

Table 2 Performance comparison on the NCI-ALMANAC dataset. Bold values indicate the best performance

	Random split			Leave-cell-out			Leave-combination out		
	RMSE	R^2	PCC	RMSE	R^2	PCC	RMSE	R^2	PCC
PermuteDDS	**43.053**	**0.527**	**0.726**	54.128	0.242	0.519	**51.58**	**0.318**	**0.569**
HypergraphSynergy	43.89	0.508	0.719	**53.398**	**0.273**	**0.538**	52.609	0.291	0.543
DeepSynergy	44.44	0.491	0.701	54.560	0.230	0.322	53.500	0.262	0.526
ComboFM	48.27	0.399	0.651	54.670	0.245	0.531	53.890	0.267	0.526
DTF	47.03	0.430	0.678	54.730	0.223	0.517	53.470	0.263	0.531
Celebi's method	47.31	0.423	0.653	53.49	0.259	0.516	55.830	0.196	0.456

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_t^i - y_p^i\right)^2}{n}} \tag{12}$$

$$\mathrm{R}^2 = 1 - \frac{\mathrm{RMSE}^2}{\mathrm{Var}\left(y_t^i\right)}, \quad \text{where}\,\mathrm{Var}\left(y_t\right) = \sum_{i=1}^{n}\left(y_t^i - \bar{y}_t\right)^2 \tag{13}$$

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}\left(y_t^i - \bar{y}_t\right)\left(y_p^i - \bar{y}_p\right)}{\sqrt{\sum_{i=1}^{n}\left(y_t^i - \bar{y}_t\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_p^i - \bar{y}_p\right)^2}} \tag{14}$$

In the equations, y_t and y_p denotes the true synergy scores and predicted synergy scores, respectively. $\bar{y}_t$ and $\bar{y}_p$ represent the mean value of the true synergy scores and predicted synergy scores, respectively.

Results

Performance comparison on cross validation

The 5-fold cross validation results of PermuteDDS and other competitive methods on O'Neil dataset are shown in Table 1. It can be easily seen that our PermuteDDS surpassed all baselines by a large margin on the random split task, e.g. 9.4% relative R^2 increase compared to the previous state-of-art method. Compared with random split task, leave-cell-out and leave-combination-out are more challenging, which test the performance of prediction models on unseen drug combinations/cell lines. On these two tasks, the performance of all methods decreased significantly compared to random split task. PermuteDDS performed slightly worse than HypergraphSynergy on the leave-cell-out task. This may be attributed to the fact that HypergraphSynergy utilized an auxiliary task that involved reconstructing cell line and drug similarity networks, which can enhance the robustness of the model on leave-cell-out task. However, this auxiliary task took all drugs and cell lines as input, resulting in the presence of unknown combinations and cell lines solely within the prediction module. Despite these considerations, the performance gap between our method and HypergraphSynergy remains quite small. As for the leave-combination-out task, PermuteDDS achieved the best results, with 19.3% relative R^2 increase and 8.1% relative PCC increase compared to HypergraphSynergy. These results highlight the superior predictive capabilities of PermuteDDS in accurately estimating drug synergy.

Table 2 shows that the performance of the models is comparatively lower on the NCI-ALMANAC dataset in comparison to the O'Neil dataset. This observation could be attributed to the fact that the NCI-ALMANAC dataset encompasses a wider range of drugs and cell lines, thereby increasing the complexity of prediction. However, PermuteDDS still achieved better performance

Table 3 Performance comparison on the independent dataset. Bold values indicate the best performance

	O'Neil			NCI-ALMANAC		
	RMSE	R2	PCC	RMSE	R2	PCC
PermuteDDS	**15.144**	**0.659**	**0.821**	**43.338**	**0.484**	0.696
HypergraphSynergy	16.710	0.585	0.788	43.730	0.474	0.693
DeepSynergy	16.840	0.578	0.765	45.325	0.435	0.670
ComboFM	16.080	0.541	0.754	46.370	0.457	0.685
DTF	16.150	0.548	0.752	49.860	0.372	**0.700**
Celebi's method	16.500	0.529	0.728	45.860	0.469	0.688

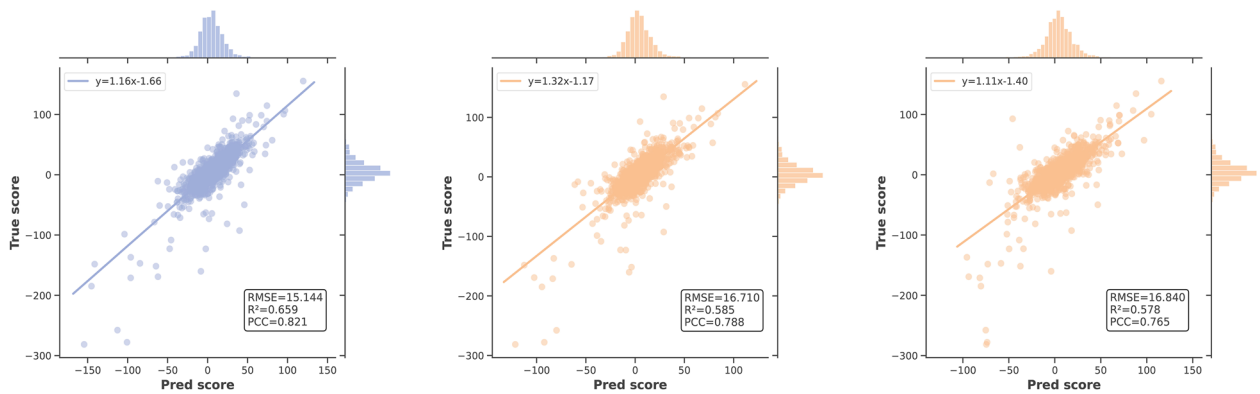

Fig. 4 Independent test results on the O'Neil dataset. From left to right: PermuteDDS, HypergraphSynergy and DeepSynergy

in most cases. Regarding the random split task, PermuteDDS demonstrated superior performance compared to other state-of-the-art approaches. It achieved a RMSE of 43.053, a R^2 of 0.527 and a PCC of 0.726. However, similar to the results on the O'Neil dataset, our method only achieved competitive results compared to other methods. However, in the leave-combination-out task, PermuteDDS outperformed all other baseline methods by a large margin, achieving the lowest RMSE of 51.58, the highest R^2 of 0.318 and the highest PCC of 0.569.

Performance evaluation on independent test

Furthermore, we assessed the generalization performance of our model by testing on independent test sets, and the results are presented in Table 3. Notably, PermuteDDS achieved the overall best performance. Specifically, on the O'Neil dataset, PermuteDDS attained the top result with RMSE, R^2, and PCC scores of 15.144, 0.659, and 0.821, respectively. While the NCI-ALMANAC dataset's complexity posed challenges for prediction across all methods, PermuteDDS still exhibited a slight superiority over previous state-of-the-art approaches with RMSE, R^2, and PCC scores of 43.338, 0.484, and 0.696, respectively. To intuitively assess differences in predictive performance across different datasets, we analyzed the distribution of true scores and predicted scores generated by the top three methods—PermuteDDS, HypergraphSynergy, and DeepSynergy. Figure 4 illustrates that the prediction results of all models on the O'Neil dataset form a well-clustered fitting line, indicative of good predictive performance. Conversely, as depicted in Fig. 5, the results on the ALMANAC datasets appear relatively dispersed, and the expansion of synergy scores (coordinate axis) range further confirms the complexity of this dataset.

Ablation study

To study the effectiveness of each inputs and each PermuteDDS units, we perform the ablation studies under random split on both datasets. As shown in Table 4, the complete PermuteDDS framework achieves the best performance on 5 of 6 evaluation. In terms of the selection of input, we designed variants with different molecular fingerprint combinations as inputs. Upon examination, it becomes evident that the removal of any of the three fingerprints results in a decline in performance, and employing only one fingerprint yields even worse results. We can infer from the results that the combination of the three different fingerprints complements each other jointly contributes to the predictive performance of

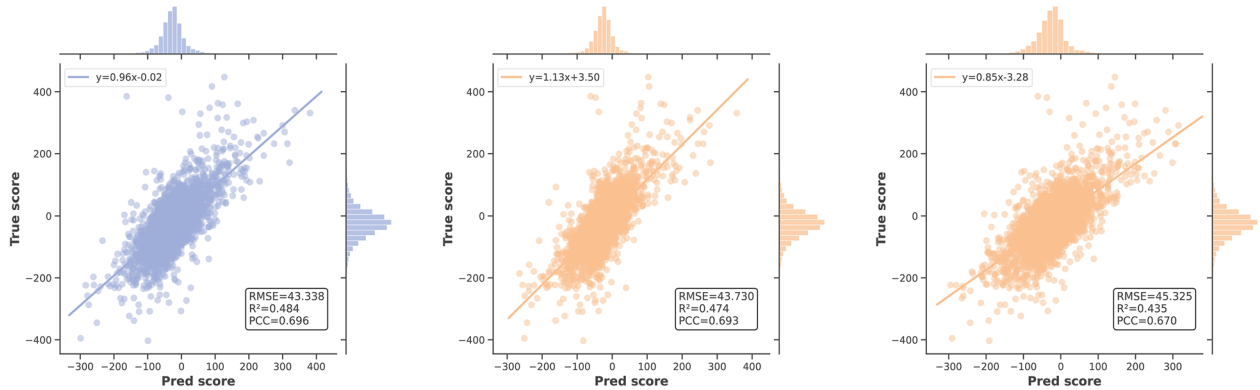

Fig. 5 Independent test results on the O'Neil dataset. From left to right: PermuteDDS, HypergraphSynergy and DeepSynergy

Table 4 Results of ablation study. Bold values indicate the best performance

	O'Neil			NCI-ALMANAC		
	RMSE	R^2	PCC	RMSE	R^2	PCC
PermuteDDS	**13.721**	**0.641**	**0.801**	**43.053**	**0.527**	0.726
w/o hashtt	13.895	0.632	0.796	44.081	0.504	0.711
w/o MAP4	13.746	0.639	0.800	43.964	0.507	0.713
w/o maccs	13.953	0.628	0.793	44.185	0.501	0.709
hashtt	14.340	0.607	0.780	46.034	0.459	0.681
maccs	14.444	0.602	0.777	46.141	0.456	0.681
MAP4	14.345	0.607	0.780	46.310	0.452	0.678
w/o Permute-MLP	14.811	0.581	0.769	44.528	0.494	0.718
PermuteDDS-L	14.433	0.602	0.777	43.138	0.525	**0.728**

PermuteDDS. Then, in terms of model design, we conducted another two variants: PermuteDDS without Permute-MLP block and PermuteDDS-L. PermuteDDS-L is the model that utilizes a feedforward network to extract features from fingerprints and cell lines. The model without Permute-MLP block demonstrated inferior performance compared to PermuteDDS, highlighting the effectiveness of this module. Moreover, CNN might not capture as much information from fingerprints and cell lines as expected, as PermuteDDS-L only performed slightly worse than PermuteDDS.

Model performance is robust against noise cell lines

In actual clinical situations, cancer cells may be mixed with normal cells. To assess the robustness of PermuteDDS, we conducted a sensitivity analysis to evaluate the stability of model performance in response to noise, following the methodology outlined in [45]. Specifically, we introduced different levels of multiplicative or additive Gaussian noise into the input cell line gene expression and mutation profiles. Subsequently, PermuteDDS was trained using these resulting noisy cell lines. The underlying assumption is that Gaussian noise injected into the data can serve as a simulation of normal cell lines. For a given gene expression (or mutation) profile X, the computation of input noisy cell lines is as follows:

$$x_{mul} = x * N(1, \sigma_{mul}) \tag{15}$$

$$x_{add} = x + N(0, \sigma_{add}) \tag{16}$$

where x_{mul} represent multiplicative noisy cell lines generated with a Gaussian distribution $N(1, \sigma_{mul})$, and x_{add} represent additive noisy cell lines generated with a Gaussian distribution $N(0, \sigma_{add})$. We adopted the same standard deviation as reported in [45]. Subsequently, we performed 5-fold cross-validation under random split across each standard deviation on both datasets. As depicted in Fig. 6A, the predicted synergy scores, obtained by training on cell lines with multiplicative Gaussian noise, exhibit consistently high correlations with those trained on the original cell lines. Remarkably, even as the magnitude of the noise increases, the stability of the model's performance is retained. Similar behavior

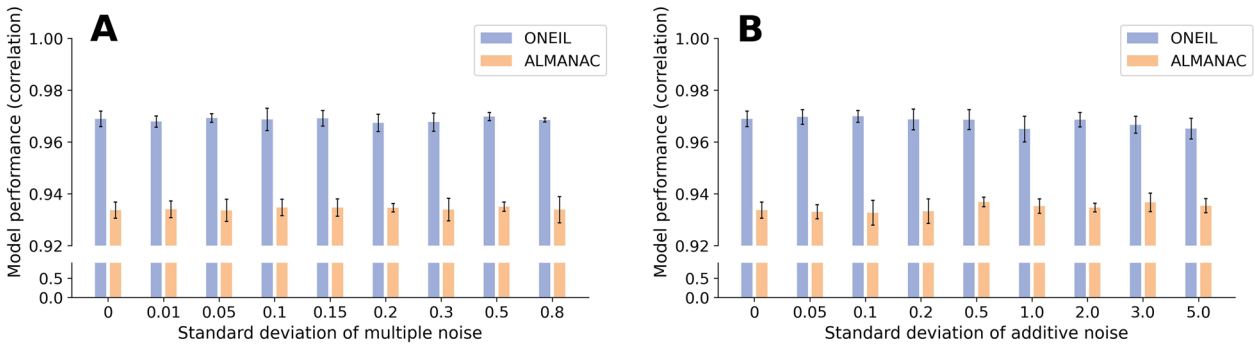

Fig. 6 Correlation between predicted synergy scores of cell lines without noise and cell lines with multiplicative noise (A) or additive noise (B)

Table 5 Performance comparison of different cell lines descriptors on the O'Neil dataset

	Random split			Leave-cell-out			Leave-combination-out		
	RMSE	R^2	PCC	RMSE	R^2	PCC	RMSE	R^2	PCC
cline-gdsc	13.721	0.641	0.801	19.668	0.243	0.522	16.152	0.501	0.709
cline-cosmic	13.602	0.647	0.805	19.928	0.236	0.508	16.122	0.503	0.71
cline-onehot	13.663	0.644	0.803	20.010	0.232	0.507	16.175	0.499	0.708

Table 6 Performance comparison of different cell lines descriptors on the NCI-ALMANAC dataset

	Random split			Leave-cell-out			Leave-combination-out		
	RMSE	R^2	PCC	RMSE	R^2	PCC	RMSE	R^2	PCC
cline-gdsc	43.053	0.527	0.726	54.128	0.242	0.519	51.58	0.318	0.569
cline-cosmic	43.247	0.522	0.723	54.83	0.218	0.515	51.303	0.325	0.575
cline-onehot	43.101	0.526	0.726	55.336	0.209	0.506	51.53	0.319	0.57

is observed when subjecting PermuteDDS to additive Gaussian noise, as illustrated in Fig. 6B. Based on these observations, we can infer that PermuteDDS demonstrates robustness in the presence of noisy data.

Cell line data study

Cell lines are known to exhibit sensitivity to variations in experimental conditions. Even identical cell lines obtained from different institutions can exhibit distinct gene expression profiles. To assess the necessity of the selected cell line descriptors, we constructed the variant of the PermuteDDS employing the cell line gene expression profiles sourced from the Cell Lines Project data in the COSMIC database [46]. Following the methodology outlined in [24], we considered only data related to 651 genes from the COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census.). Moreover, we created an another variant utilizing a simple one-hot encoding as the input cell line descriptor, serving as a baseline for comparison. For the leave-cell-out task, the cell lines within the test set, which are unseen during training, are encoded as simple zero vectors. The results on the O'Neil and NCI-ALMANAC datasets are presented in Table 5 and Table 6, respectively. The term 'cline-gdsc' denotes the original PermuteDDS, whereas 'cline-cosmic' and 'cline-onehot' refer to the variants employing COSMIC and one-hot cell line descriptors, respectively. The outcomes derived from employing different cell lines descriptors consistently exhibit similar performance across all cross-validation scenarios. This suggests that our method demonstrates insensitivity towards the choice of these cell-line descriptors, as long as the representation method adequately represents and distinguishes different cell lines. It's crucial to highlight that all variants exhibited poor performance under the leave-cell-out task. Even upon encoding an unseen cell line with a zero vector, no significant differences were observed in comparison to the utilization of gene expression profiles. Indeed, none of the baseline methods demonstrate the capability to achieve satisfactory results on

this task (Tables 1 and 2), underscoring the considerable challenge inherent in its execution.

Furthermore, we conducted similar experiments using one-hot encoding as cell line descriptors to construct the other two baseline methods, DeepSynergy and HypergraphSynergy. As presented in Additional file 1: Tables S1 and S2, DeepSynergy exhibited a similar phenomenon to PermuteDDS, wherein employing one-hot encoding as a cell line descriptor yielded similar results to the original. However, replacing the cell line descriptor with one-hot encoding leads to a significant decline in the performance of HypergraphSynergy. This may be attributed to the fact that HypergraphSynergy incorporates the reconstruction of cell line similarities as part of its optimization objective. Since the similarity between cell lines encoded using one-hot encoding is uniformly zero, this reconstruction process loses its significance. These observations reveal the inconsistent sensitivity among different methods to the selection of cell line descriptors, depending on the preprocessing approach applied to the cell lines.

Predicting novel synergistic combinations

In this section, we employed PermuteDDS to predict novel synergistic drug combinations that had not been previously tested. We utilized all measured trios of drug pairs and cell lines to train PermuteDDS and subsequently made predictions for unmeasured trios using the NCI-ALMANAC dataset. We focused on drug combinations with predicted scores close to 1 (see GitHub link[3].). We further conducted a nonexhaustive literature search, which revealed that six of the predicted drug combinations were consistent with observations from previous studies. For example, Dasatinib and Gefitinib combination presented a cell-specific cytotoxic synergistic effect in human ovarian cell line OVCAR-3 and IGROV-1 [47]. According to the trials of Dolfi et al., combination of fulvestrant and doxorubicin can enhance the sensitivity of breast cancer cell line T47D to these cytotoxic agents [48]. We believe that there are other predicted drug pairs that hold the potential of being promising combinations, which require further validation.

Discussion

From the results, while PermuteDDS has demonstrated outstanding performance, we noticed that its performance on the leave-cell-out and leave-combination-out tasks is limited. The same situation occurred on other baselines, where R^2 and PCC scores are consistently below 0.5 and 0.7, respectively. These scores indicate that the predicted results of the model are almost meaningless for these tasks. The reason for this limitation may be attributed to the distribution shift between the training and test sets, caused by the disparity in drug combinations and cell lines between these sets. The problem is expected to be solved by learning the invariance between different drugs and cell lines [49, 50]. Our future work is to explore a more robust model for these leave-out cross validation tasks.

In 3.3, we deduced that the combined use of the three different fingerprints significantly contributed to the predictive performance of PermuteDDS. This observation is likely because these fingerprints provide distinct descriptions of molecules from various perspectives. HashTT fingerprint is a type of path-based fingerprint that incorporates the topological information of molecules [34, 51]. MACCS keys, on the other hand, generate bit strings based on the presence or absence of specific substructures or features, thus enabling the capture of structural information. MAP4 is a novel fingerprint that combines the atom-pair approach with circular substructures, allowing it to encode both molecular shape and chemical information simultaneously [35]. Thus, HashTT provides valuable topological information, while MACCS offers essential structural information. Subsequently, MAP4 supplements the chemical properties from the atomic perspective. The fusion of these distinct pieces of information results in a comprehensive and detailed description of the molecules.

In "Cell line data study" section, we conducted experiments to investigate the significance of cell line descriptors. The results of these experiments suggest that different methods have varying sensitivities to the choice of cell line descriptors, indicating the limitations of one-hot encoding for certain methods. Moreover, the utilization of different cell line descriptors all resulted in poor performance under the leave-cell-out task. Several other related studies have also demonstrated poor generalization performance on this task [52–54]. Nevertheless, we maintain the conviction that continued research in fields such as pharmacology, pharmacokinetics, toxicology, and genetic heterogeneity, alongside the development of novel computational methods, holds the potential to swiftly surmount these challenges.

Conclusions

In conclusion, we proposed PermuteDDS, a novel model designed to predict potential synergistic drug combinations for cancer treatment. PermuteDDS establishes a unified framework that incorporates diverse types of information, including topological structure and chemical properties of drugs, cellular gene expressions and gene mutations. These different

[3] https://github.com/littlewei-lazy/PermuteDDS/blob/main/Data/novel_synergy_predict.csv

data are effectively fused using the FSN architecture to capture the complex interactions between drug pairs and cell lines. PermuteDDS exhibits robust predictive capabilities on two benchmark datasets through comparison with other competitive methods. However, there remain certain limitations that have been previously discussed. Our future work is to explore a more robust model for leave-out cross validation tasks.

Abbreviations

HTS	High-throughput screening
SVM	Support vector machine
XGBoost	Extreme gradient boosting
SMILES	Simplified molecular-input line-entry system
CNN	Convolutional neural network
GCN	Graph convolutional network
HashTT	Hashed topological-torsion
MAP4	MinHashed atom-pair fingerprint up to a diameter of four bonds
MACCS	Molecular ACCess System
DDS	Drug-drug synergy
GDSC	Genomics of Drug Sensitivity in Cancer
FSNs	Fingerprint specific networks
HGNN	Hypergraph neural net
PPI	Protein-protein-interaction
Conv1D	One-dimensional convolutional neural network
MSE	Mean square error
RMSE	Root man squared error
R^2	Coefficient of determination
PCC	Pearson's Correlation Coefficient

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-024-00839-8.

Additional file 1: Table S1. Performance comparison of one-hot cell line descriptors with different baselines methods on O'Neil dataset. **Table S2.** Performance comparison of one-hot cell line descriptors with different baselines methods on NCI-ALMANAC dataset.

Acknowledgements

Our deepest gratitude goes to the anonymous reviewers for their careful work and insightful suggestions, which will undoubtedly enhance the quality of this paper.

Author contributions

X.Z. prepared the data, performed the experiments and wrote the manuscript, J.X. analyzed the results and revised the manuscript, Y.S. and M.X. participated in revising the manuscript, J.H. and X.L. contributed to the interpretation and revising of the manuscript, K.C. organized the database, J.W. and Y.L. supervised the research activity planning and execution.

Funding

This research was funded by National Natural Science Foundation of China (NSFC, Grant No. 62,271,174, 62,202,373 and 62,102,191); the Industry Prospecting and Common Key Technology Key Projects of Jiangsu Province Science and Technology Department (Grant no.BE2020721); Industrial Chain Collaborative Innovation Project of Ministry of Industry and Information Technology "Multi-modal Medical Data Intelligent Management Software for the New Generation of Information Technology"(TC210804V); the Industrial and Information Industry Transformation and Upgrading Special Fund of Jiangsu Province in 2021 (Grant no. [2021]92); the Key Project of Smart Jiangsu in 2020 (Grant no. [2021]1); Jiangsu Province Engineering Research Center of Big Data Application in Chronic Disease and Intelligent Health Service (Grant no. [2020]1460).

Data availibility

Lists the following:

- Project name: PermuteDDS
- Project home page: https://github.com/littlewei-lazy/PermuteDDS
- Operating system(s): Linux, Windows
- Programming language: Python
- Other requirements: Python 3.7 or higher, PyTorch 1.11 or higher

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no Conflict of interest.

Received: 19 September 2023 Accepted: 3 April 2024
Published online: 15 April 2024

References

1. Lopez JS, Banerji U (2017) Combine and conquer: challenges for targeted therapy combinations in early phase trials. Nat Rev Clin Oncol 14(1):57–66
2. Al-Lazikani B, Banerji U, Workman P (2012) Combinatorial drug therapy for cancer in the post-genomic era. Nat Biotechnol 30(7):679–692
3. Liu J, Gefen O, Ronin I, Bar-Meir M, Balaban NQ (2020) Effect of tolerance on the evolution of antibiotic resistance under drug combinations. Science 367(6474):200–204
4. Azam F, Vazquez A (2021) Trends in Phase II Trials for Cancer Therapies. Cancers 2021, 13, 178. s Note: MDPI stays neu-tral with regard to jurisdictional clai-ms in ..
5. Li P, Huang C, Fu Y, Wang J, Wu Z, Ru J, Zheng C, Guo Z, Chen X, Zhou W et al (2015) Large-scale exploration and analysis of drug combinations. Bioinformatics 31(12):2007–2016
6. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1(11):882–894
7. Ferreira D, Adega F, Chaves R (2013) The importance of cancer cell lines as in vitro models in cancer methylome analysis and anticancer drugs testing. Oncogenomics 1:139–166
8. Morris MK, Clarke DC, Osimiri LC, Lauffenburger DA (2016) Systematic analysis of quantitative logic model ensembles predicts drug combination effects on cell signaling networks. CPT 5(10):544–553
9. Preuer K, Lewis RP, Hochreiter S, Bender A, Bulusu KC, Klambauer G (2018) Deepsynergy: predicting anti-cancer drug synergy with deep learning. Bioinformatics 34(9):1538–1546
10. Breiman L (2001) Random forests. Mach Learn 45:5–32
11. Noble WS (2006) What is a support vector machine? Nat Biotechnol 24(12):1565–1567
12. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T et al (2015) Xgboost: extreme gradient boosting. R package version 0.4-2 1(4):1–4
13. Kumar V, Dogra N (2022) A comprehensive review on deep synergistic drug prediction techniques for cancer. Arch Comput Methods Eng 29(3):1443–1461
14. Jeon M, Kim S, Park S, Lee H, Kang J (2018) In silico drug combination discovery for personalized cancer therapy. BMC Syst Biol 12(2):59–67
15. He L, Tang J, Andersson EI, Timonen S, Koschmieder S, Wennerberg K, Mustjoki S, Aittokallio T (2018) Patient-customized drug combination prediction and testing for t-cell prolymphocytic leukemia patients. Can Res 78(9):2407–2418
16. Kuru HI, Tastan O, Cicek AE (2021) Matchmaker: a deep learning framework for drug synergy prediction. IEEE/ACM Trans Comput Biol Bioinf 19(4):2334–2344
17. Cao D-S, Xu Q-S, Hu Q-N, Liang Y-Z (2013) Chemopy: freely available python package for computational biology and chemoinformatics. Bioinformatics 29(8):1092–1094

18. Lin W, Wu L, Zhang Y, Wen Y, Yan B, Dai C, Liu K, He S, Bo X (2022) An enhanced cascade-based deep forest model for drug combination prediction. Brief Bioinform 23(2):562
19. Hosseini S-R, Zhou X (2023) Ccsynergy: an integrative deep-learning framework enabling context-aware prediction of anti-cancer drug synergy. Brief Bioinform 24(1):588
20. Weininger D (1988) Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. J Chem Inf Comput Sci 28(1):31–36
21. Kim Y, Zheng S, Tang J, Jim Zheng W, Li Z, Jiang X (2021) Anticancer drug synergy prediction in understudied tissues using transfer learning. J Am Med Inform Assoc 28(1):42–51
22. Yu L, Su Y, Liu Y, Zeng X (2021) Review of unsupervised pretraining strategies for molecules representation. Brief Funct Genomics 20(5):323–332
23. Wang J, Liu X, Shen S, Deng L, Liu H (2022) Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. Brief Bioinform 23(1):390
24. Liu X, Song C, Liu S, Li M, Zhou X, Zhang W (2022) Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction. Bioinformatics 38(20):4782–4789
25. O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, Li J, Kral A, Lejnine S, Loboda A et al (2016) An unbiased oncology compound screen to identify novel combination strategies. Mol Cancer Ther 15(6):1155–1162. https://doi.org/10.1158/1535-7163.MCT-15-0843
26. Holbeck SL, Camalier R, Crowell JA, Govindharajulu JP, Hollingshead M, Anderson LW, Polley E, Rubinstein L, Srivastava A, Wilsker D et al (2017) The national cancer institute almanac: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. Can Res 77(13):3564–3576
27. Loewe S (1953) The problem of synergism and antagonism of combined drugs. Arzneimittelforschung 3(6):285–290
28. Bliss CI, The toxicity of poisons applied jointly. Ann Appl Biol 26(3):585–615. https://doi.org/10.1111/j.1744-7348.1939.tb06990.x
29. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B et al (2019) Pubchem 2019 update: improved access to chemical data. Nucleic Acids Res 47(D1):1102–1109
30. Landrum G et al (2013) Rdkit: a software suite for cheminformatics, computational chemistry, and predictive modeling. Greg Landrum 8:31
31. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Gonçalves E, Barthorpe S, Lightfoot H et al (2016) A landscape of pharmacogenomic interactions in cancer. Cell 166(3):740–754
32. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR et al (2012) Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 41(D1):955–961
33. Cheng L, Li L (2016) Systematic quality control analysis of lincs data. CPT 5(11):588–598
34. Nilakantan R, Bauman N, Dixon JS, Venkataraghavan R (1987) Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. J Chem Inf Comput Sci 27(2):82–85
35. Capecchi A, Probst D, Reymond J-L (2020) One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. J Cheminf 12(1):1–15
36. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of mdl keys for use in drug discovery. J Chem Inf Comput Sci 42(6):1273–1280
37. Hendrycks D, Gimpel K (2016) Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415
38. Tolstikhin IO, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, Yung J, Steiner A, Keysers D, Uszkoreit J et al (2021) Mlp-mixer: an all-mlp architecture for vision. Adv Neural Inf Process Syst 34:24261–24272
39. Hou Q, Jiang Z, Yuan L, Cheng M-M, Yan S, Feng J (2022) Vision permutator: a permutable MLP-like architecture for visual recognition. IEEE Trans Pattern Anal Mach Intell 45(1):1328–1334
40. Ramachandran P, Zoph B, Le QV (2017) Searching for activation functions. arXiv preprint arXiv:1710.05941
41. Ba JL, Kiros JR, Hinton GE (2016) Layer normalization. arXiv preprint arXiv:1607.06450
42. Sun Z, Huang S, Jiang P, Hu P (2020) Dtf: deep tensor factorization for predicting anticancer drug synergy. Bioinformatics 36(16):4483–4489
43. Julkunen H, Cichonska A, Gautam P, Szedmak S, Douat J, Pahikkala T, Aittokallio T, Rousu J (2020) Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. Nat Commun 11(1):6136
44. Celebi R, Bear Don't Walk O, Movva R, Alpsoy S, Dumontier M (2019) In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. Sci Rep 9(1), 1–10
45. Yuan B, Shen C, Luna A, Korkut A, Marks DS, Ingraham J, Sander C (2021) Cellbox: interpretable machine learning for perturbation biology with application to the design of cancer combination therapy. Cell Syst 12(2):128–140
46. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S et al (2015) Cosmic: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 43(D1):805–811
47. Thibault B, Jean-Claude B (2017) Dasatinib+ gefitinib, a non platinum-based combination with enhanced growth inhibitory, anti-migratory and anti-invasive potency against human ovarian cancer cells. Journal of ovarian research 10:1–12
48. Dolfi SC, Jäger AV, Medina DJ, Haffty BG, Yang J-M, Hirshfield KM (2014) Fulvestrant treatment alters mdm2 protein turnover and sensitivity of human breast carcinoma cells to chemotherapeutic drugs. Cancer Lett 350(1–2):52–60
49. Shen Z, Liu J, He Y, Zhang X, Xu R, Yu H, Cui P (2021) Towards out-of-distribution generalization: a survey. arXiv preprint arXiv:2108.13624
50. Yang N, Zeng K, Wu Q, Jia X, Yan J (2022) Learning substructure invariance for out-of-distribution molecular representations. Adv Neural Inf Process Syst 35:12964–12978
51. O'Boyle NM, Sayle RA (2016) Comparing structural fingerprints using a literature-based similarity benchmark. J Cheminf 8(1):1–14
52. Wang X, Zhu H, Chen D, Yu Y, Liu Q, Liu Q (2023) A complete graph-based approach with multi-task learning for predicting synergistic drug combinations. Bioinformatics 39(6):351
53. Zhang P, Tu S (2023) MGAE-DC: Predicting the synergistic effects of drug combinations through multi-channel graph autoencoders. PLoS Comput Biol 19(3):1010951
54. Preto AJ, Matos-Filipe P, Mourão J, Moreira IS (2022) Synpred: prediction of drug combination effects in cancer using different synergy metrics and ensemble learning. GigaScience 11:087

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.