Interdisciplinary Sciences: Computational Life Sciences (2021) 13:299–311
https://doi.org/10.1007/s12539-021-00422-x

ORIGINAL RESEARCH ARTICLE

An In Silico Method for Predicting Drug Synergy Based on Multitask Learning

Xin Chen[1] · Lingyun Luo[1,2] · Cong Shen[3] · Pingjian Ding[1,2] · Jiawei Luo[3]

Received: 6 November 2020 / Revised: 29 January 2021 / Accepted: 7 February 2021 / Published online: 21 February 2021

Abstract
To make better use of all kinds of knowledge to predict drug synergy, it is crucial to successfully establish a drug synergy prediction model and leverage the reconstruction of sparse known drug targets. Therefore, we present an in silico method that predicts the synergy scores of drug pairs based on multitask learning (DSML) that could fuse drug targets, protein–protein interactions, anatomical therapeutic chemical codes, a priori knowledge of drug combinations. To simultaneously reconstruct drug–target protein interactions and synergistic drug combinations, DSML benefits indirectly from the associations with relation through proteins. In cross-validation experiments, DSML improved the ability to predict drug synergy. Moreover, the reconstruction of drug–target interactions and the incorporation of multisource knowledge significantly improved drug combination predictions by a large margin. The potential drug combinations predicted by DSML demonstrate its ability to predict drug synergy.

Keywords Drug synergy · Multitask learning · Drug–target interaction · In silico technology

1 Introduction

Combined drug therapy involves giving two or more drugs to patients either simultaneously or sequentially, and such drug combinations are particularly important treatments for battling cancer and acquired immune deficiency syndrome (AIDS). The traditional "one drug, one target" therapy has limited therapeutic effects due to side effects and drug resistance [1, 2]; co-administered drugs may increase therapeutic efficacy and reduce unnecessary off-target effects [3–5]. As commercialization has progressed, the U.S. Food and Drug Administration (FDA) has approved an increasing number of fixed-dose combinations for the treatment of complex diseases, such as neoplasms, AIDS, and type II diabetes. In January 2014, the FDA approved the first combination of drugs to combat melanoma with BRAF V600E or V600K mutations [6].

At the pharmacological level, relationships among individual drugs can be divided into three categories of action: antagonistic, additive, and synergistic, and the effect of a drug combination is equal to, less than or greater than the sum of each drug [7]. Synergistic effects not only help reduce dosages while maintaining equal levels of efficacy but also delay the development of drug resistance, which is an ideal goal in the field of drug discovery [5, 7, 8]. Nevertheless, developing a promising method for discovering synergistic drug combinations to better overcome complex diseases remains a challenge.

Using the traditional 'case-by-case' method, it is expensive and time-consuming with to identify drug combinations from the vast pool of potential drug pairs. Therefore, due to the enormous screening space, some in silico methods have recently been proposed to improve clinical trials [9, 10]. Various patterns, including pharmacological features [11], and network topological features [12–14], are enriched in a large number of effective drug combinations; these patterns were used to build a statistical learning model. Then, computational models to calculate a synergy score can effectively screen drug combinations from the many potential drug combinations based on numerical features extracted from

✉ Pingjian Ding
dpj@usc.edu.cn

[1] School of Computer Science, University of South China, Hengyang 421001, Hunan, China

[2] Hunan Medical Big Data International Sci.&Tech. Innovation Cooperation Base, Hengyang 421000, Hunan, China

[3] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, Hunan, China

these distinguished features. These computational models fall into two categories [15]: (1) hypothesis-driven and (2) data-driven. In recent years, several efforts have been made to construct hypothesis-driven models for synergistic drug combination prediction [16]. Network-based algorithms were used to develop prediction models using the hypotheses. For example, NIMS [16] performed multi-view analysis through network community analysis based on the hypothesis that the phenotypic similarity of drug–target-related diseases might be combined with target proximity in the drug–target network. CDA [17] used a drug-induced gene expression profile to analyze the correlations between drug-induced pathways and disease-related pathways. DIGRE [18] combined the similarity of expression profiles of drug genes and drug response curves to predict a synergistic score for drug pairs. SynGen [19] estimated drug synergy by performing enrichment analysis on the expression profiles of drug-induced genes and disease-related genes. DCPred [20] utilized a set-based statistical method to assign drug synergy scores on a drug–drug cocktail network. The network-based computational biology method [21] investigated drug–target proteins in genetic interaction networks and drug-related pathways to explore effective drug combinations; this work improves our understanding of the potential rules for finding synergistic drug combinations. Considering that priori knowledge contributes to discovering synergistic drug combinations, some computational approaches have been developed based on the features of known synergistic drug combinations; these are called data-driven models [15]. PDC-SGB [22] applied a stochastic gradient boosting algorithm to calculate a drug synergy score by incorporating drug structure similarity, target gene similarity, ATC similarity and drug combinations. RACS [23] proposed a semi-supervised sequencing model to predict drug synergy by considering the distinguishable features of effective drug combinations. NLLSS [24] separately analyzed results from different spaces and then combined these results using a simple weighted average method. Recently, a new ensemble prediction framework, named EPSDC [25], was proposed to integrate information from multiple sources to prioritize potential synergistic drug combinations. However, all these computational methods predict drug synergy based on sparse drug–target interactions, which may affect their prediction performances to some extent.

Overall, with the emergence of computational network pharmacology [26], the exploration of multidrug pharmacology has made substantial progress and improved the effectiveness of targeted therapies. However, drug combination prediction results could be improved by reconstructing drug targets. Although computational models have been proposed from multiple aspects for predicting drug targets [27, 28], these methods all consider drug–target prediction and drug combination discovery as two separate learning tasks in different computational models. Thus, we integrated drug–target prediction [29, 30] and drug combination prediction [15] into a unified framework, namely the DSML algorithm, which can accurately and effectively predict drug synergy. First, we calculate drug similarity based on values provided by the anatomical therapeutic chemistry (ATC) coding system, and we calculate protein similarity using paths from the protein–protein interaction (PPI) network. Then, we defined graph regularization and used the dual least squares method to establish a new objective function for a unified framework to predict both drug targets and drug synergies that can incorporate both multi-type drug and protein networks. In cross-validation experiments, on the DCDB database [31], we compared DSML with existing drug synergy prediction methods and evaluated the effectiveness of incorporating various source knowledge. Finally, to further demonstrate our framework's ability to calculate a drug synergy score, we present some drug pairs with the highest synergy score as calculated by DSML.

2 Materials and Methods

The DSML model in this study reconstructs drug–target interactions to increase the prediction performance when scoring drug synergy through a unified framework. The schematic representation of the DSML is described in Fig. 1. In DSML, we first calculate drug similarity by applying the Jaccard index to the drug ATC code. Then, a path-based algorithm is designed to quantify protein similarity using the topology of the PPI network. Finally, we design a multitask learning algorithm to improve the drug synergy scoring performance by reconstructing drug–target interactions, which can incorporate known drug combinations, protein topological similarity, and drug anatomical therapeutic similarity. Furthermore, all the notations defined in DSML are provided and summarized in Table 1.

2.1 A Priori Drug Combination Knowledge

The enriched patterns hidden in priori knowledge of drug combinations, such as network topological features [12] and pharmacological features [11], can be used to build statistical learning models to predict drug synergy. Therefore, incorporating priori knowledge of drug combinations into models can contribute to screening drug combinations from many drug pairs. In this study, all the effective drug combinations were provided by DCDB 2.0 database, which forms a useful basis for computational modeling to predict synergistic scores between drugs and is freely available from http://www/cls.zju.edu.cn/dcdb/ [31].

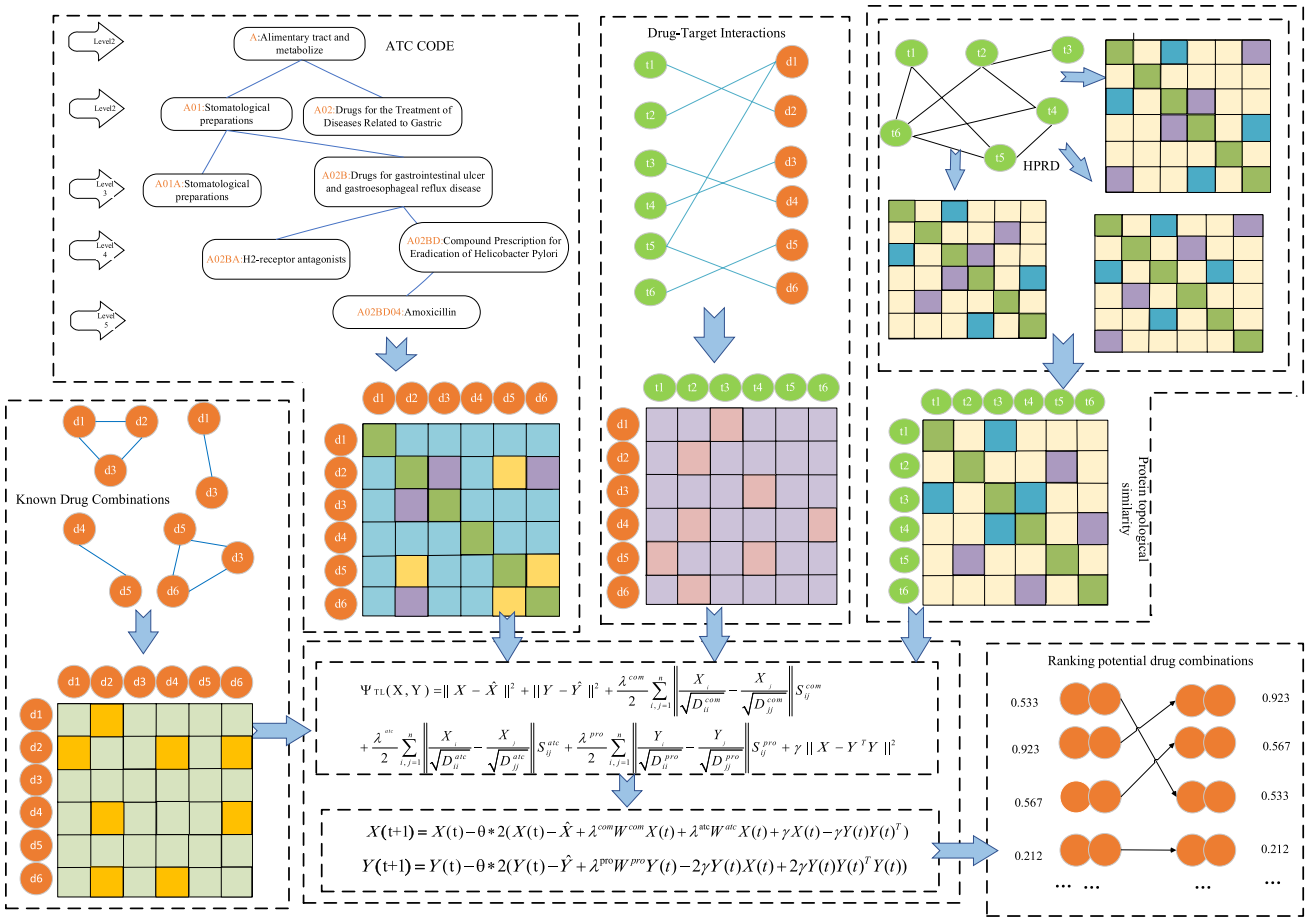

Fig. 1 Scheme to predict the synergy of drug pairs based on multitask learning

Table 1 Notations

Notation	Definition
m,n	Number of proteins, and drugs
$\hat{X} \in \mathbb{R}_{\geq 0}^{n*n}$	A priori knowledge of drug combinations used in training
$X \in \mathbb{R}_{\geq 0}^{n*n}$	Drug synergy for learning
$\hat{Y} \in \mathbb{R}_{\geq 0}^{n*m}$	Known drug–target interactions for training
$Y \in \mathbb{R}_{\geq 0}^{n*m}$	Confidence score of drug target for learning
$S^{atc} \in \mathbb{R}_{\geq 0}^{n*n}$	The drug similarity matrix
$S^{pro} \in \mathbb{R}_{\geq 0}^{m*m}$	The protein similarity matrix
$S^{com} \in \mathbb{R}_{\geq 0}^{n*n}$	The drug synergy matrix
I^{n*n}	The identity matrix of size n*n
I^{m*m}	The identity matrix of size m*m
$W^{atc} \in \mathbb{R}_{\geq 0}^{n*n}$	$W^{atc} = I^{n*n} - D^{atc^{-1/2}} S^{atc} D^{atc^{-1/2}}$, where D^{atc} is a diagonal matrix, and $D_{ii}^{atc} = \sum_{j=1} S_{ij}^{atc}$
$W^{pro} \in \mathbb{R}_{\geq 0}^{m*m}$	$W^{pro} = I^{m*m} - D^{pro^{-1/2}} S^{pro} D^{pro^{-1/2}}$, where D^{pro} is a diagonal matrix, and $D_{ii}^{pro} = \sum_{j=1} S_{ij}^{pro}$
$W^{com} \in \mathbb{R}_{\geq 0}^{n*n}$	$W^{com} = I^{n*n} - D^{com^{-1/2}} S^{com} D^{com^{-1/2}}$, where D^{com} is a diagonal matrix, and $D_{ii}^{com} = \sum_{j=1} S_{ij}^{com}$

2.2 Drug Anatomical Therapeutic Similarity Calculation

A previous study of 59 effective drug combinations showed that synergistic agents are often classified into the same anatomical and therapeutic groups [11]. That is, the pharmacological characteristics of drugs can be used to distinguish effective drug combinations from a massive search space. In addition, the drug ATC classification system code reflects drug similarity, and this concept is widely used to discover new drug targets and drug combinations [32]. We utilized the Jaccard similarity method to calculate drug similarity based on the ATC codes provided by DrugBank [33]. Let $ATC_k(d_1)$ and $ATC_k(d_2)$ denote the k-th layer of the ATC

code of drugs d_1 and d_2, respectively. Then, the anatomical therapeutic similarity between drugs d_1 and d_2 is defined as follows:

$$S^{atc}(d_1, d_2) = \frac{1}{3}\sum_{k=1}^{3}\frac{|ATC_k(d_1) \cap ATC_k(d_2)|}{|ATC_k(d_1) \cup ATC_k(d_2)|}.$$

This study uses only the ATC code in the first three layers to measure drug similarity, because there is no similarity between most drugs in the fourth and fifth layers.

2.3 Protein-Related Data Acquisition and Protein Similarity Calculation

Information about the mechanism of action is another good approach for discovering synergistic drug combinations. As reported in a previous study [23], most synergistic agent targets can be reached by traversing two to four edges in the PPI network. Thus, we collected highly effective drug–target interactions from the dataset DrugBank 5.0, which is freely accessible from www.drugbank.ca [33], and the experimentally supported Homo sapiens protein–protein interactions from the HPRD database (freely accessible from http://www.hprd.org/) [34]. Protein topological similarity can be used to detect protein complexes [35, 36] and potential protein–protein interactions [37]. To date, many computational methods have been proposed to measure protein topological similarity using PPI networks, which can be mainly divided into two groups: neighbor-based and distance-based. The hypothesis of the neighbor-based approach is that two proteins are more similar when they share more neighbors [38]. The distance-based approach considers all alternative paths to measure the topological similarity of proteins in the PPI network. Lei et al. [39] developed a novel algorithm based on random walks to calculate protein topological similarity using a binary PPI network; this algorithm reduced the influences of noise of PPI network on the clustering analysis performance. To coincide with the hypothesis that numerous paths of length from 2 to 4 exist between most synergistic agent targets in the PPI network, as reported in a previous study [23], we designed a protein similarity measurement that assigns higher topological similarity to proteins connected by paths with lengths between 2 and 4. It is important to note that all proteins are considered to have self-relationships in this study. Motivated by Dice's coefficient, we developed a novel protein topological similarity measurement method based on the specific path length. The intuition behind this approach is that two similar proteins should not only be strongly connected by paths with lengths of 2–4, but should also share comparable visibility. Particularly, for special length-L paths, we defined the relatedness between proteins t_1 and t_2 as follows:

$$S^{pro-L}(t_1, t_2) = \frac{2 * P^L(t_1, t_2)}{P^L(t_1, t_1) + P^L(t_2, t_2)},$$

where $P^L(t_1, t_2)$ denotes the number of paths with length L connecting proteins t_1 and t_2, $P^L(t_1, t_1)$ denotes those connecting proteins t_1 and t_1, and $P^L(t_2, t_2)$ denotes those connecting proteins t_2 and t_2. Finally, $S^{pro-L}(t_2, t_2)$ for different lengths between 2 and 4 are combined by an averaging operation as the protein similarity $S^{pro}(t_1, t_2)$:

$$S^{pro}(t_1, t_2) = \frac{\sum_{L\in\{2,3,4\}} S^{pro-L}(t_1, t_2)}{3}.$$

2.4 Multitask Learning Design

We first introduce a base model that uses the least squares method, several variations of which have been widely applied to discover various relationships between biological entities [24, 40, 41]. To obtain the synergy score matrix X based on the known synergistic drug combination $\widehat{X}$, the objective function is defined as $\psi(X) = \sum_{i=1}^{n} \|X_i - \widehat{X}_i\|^2 = \|X - \widehat{X}\|^2$, where X_i is the i-th column of the synergy score matrix X, and $\widehat{X}_i$ is the i-th column of the known effective combinations $\widehat{X}$. Furthermore, to implement multitask learning involving both drug combination prediction and drug–target inference into a unified model, we developed a dual least squares approach by extending the base model to two fitting terms, leading to the following objective function $\psi_D(X) = \|X - \widehat{X}\|^2 + \|Y - \widehat{Y}\|^2$, where $\widehat{X}$ and $\widehat{Y}$ represent the known effective combinations and drug–target associations, respectively, and X and Y denote the synergy scores and the confidence scores between the drug and target, respectively. In the above objective function, the first and the second terms impose the fitting of a priori knowledge of drug combinations and drug–target interactions, respectively. This dual least squares introduces multitask learning because drug synergy and drug target are predicted simultaneously.

In addition, to be consistent with the data distribution of each type of knowledge, we refined the smoothness of the space using the graph regularization method. In particular, the inherent geometry of the data in each knowledge base can be preserved using the graph regularization method based on the manifold hypothesis that data points can be sampled geometrically, which has been reported in many studies [42–45]. The geometric structures of the known drug combination network, drug similarity network and protein similarity network are preserved in the following graph regularization terms:

$$\psi_{GEN}(\mathrm{X},\mathrm{Y}) = \frac{\lambda^{com}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{com}}} - \frac{X_j}{\sqrt{D_{jj}^{com}}}\right\|^2 S_{ij}^{com} + \frac{\lambda^{atc}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{atc}}} - \frac{X_j}{\sqrt{D_{jj}^{atc}}}\right\|^2 S_{ij}^{atc} + \frac{\lambda^{pro}}{2}\sum_{i,j=1}^{m}\left\|\frac{Y_i}{\sqrt{D_{ii}^{gene}}} - \frac{Y_j}{\sqrt{D_{jj}^{gene}}}\right\|^2 S_{ij}^{pro}$$

$$= \psi_{GEN}^{com}(\mathrm{X}) + \psi_{GEN}^{atc}(\mathrm{X}) + \psi_{GEN}^{pro}(\mathrm{Y}),$$

where $\lambda^{com} \in (0,1)$, $\lambda^{atc} \in (0,1)$, and $\lambda^{pro} \in (0,1)$ respectively, denote the weights of the drug combination knowledge, drug similarity knowledge, and protein similarity knowledge. In the above objective function, these three regularization terms, $\psi_{GEN}^{com}(\mathrm{X}) = \frac{\lambda^{com}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{com}}} - \frac{X_j}{\sqrt{D_{jj}^{com}}}\right\|^2 S_{ij}^{com}$, $\psi_{GEN}^{atc}(\mathrm{X}) = \frac{\lambda^{atc}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{atc}}} - \frac{X_j}{\sqrt{D_{jj}^{atc}}}\right\|^2 S_{ij}^{atc}$, and $\psi_{GEN}^{pro}(\mathrm{Y}) = \frac{\lambda^{pro}}{2}\sum_{i,j=1}^{m}\left\|\frac{Y_i}{\sqrt{D_{ii}^{gene}}} - \frac{Y_j}{\sqrt{D_{jj}^{gene}}}\right\|^2 S_{ij}^{pro}$, are three smooth constraints corresponding to drug combination knowledge, drug similarity knowledge, and protein similarity knowledge, respectively.

The in silico model based on multitask learning is introduced to simultaneously learn drug synergy and drug–target interaction, which enables multitask learning by combining the dual least squares method and the graph regularization terms of the various knowledge types. Thus, the objective function to jointly learn X and Y can be formulated as follows:

$$\psi_{TL}(\mathrm{X},\mathrm{Y}) = \|X - \hat{X}\|^2 + \|Y - \hat{Y}\|^2 + \frac{\lambda^{com}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{com}}} - \frac{X_j}{\sqrt{D_{jj}^{com}}}\right\|^2 S_{ij}^{com} + \frac{\lambda^{atc}}{2}\sum_{i,j=1}^{n}\left\|\frac{X_i}{\sqrt{D_{ii}^{atc}}} - \frac{X_j}{\sqrt{D_{jj}^{atc}}}\right\|^2 S_{ij}^{atc} + \frac{\lambda^{pro}}{2}\sum_{i,j=1}^{m}\left\|\frac{Y_i}{\sqrt{D_{ii}^{pro}}} - \frac{Y_j}{\sqrt{D_{jj}^{pro}}}\right\|^2 S_{ij}^{pro} + \gamma\|X - Y^T Y\|^2$$

$$= \psi_D(\mathrm{X}) + \psi_{GEN}(\mathrm{X},\mathrm{Y}) + \gamma\|X - Y^T Y\|^2,$$

where $\gamma \geq 0$ is a regularization parameter (set to 0.001), and the additional cost term $\|X - Y^T Y\|^2$ is used to transfer the predictive drug target to discover drug combinations, but also captures common target proteins of some synergistic agents, as reported in a previous study [8]. The proposed computational model is a transductive learning method that enables the synergistic score of unlabeled samples to be predicted without having to build a map function [46, 47]. Unlike inductive learning, which builds a map function based only on all the labeled data, DSML can simultaneously accept all data as input, including both labeled and unlabeled samples, and it predicts the scores of unlabeled samples. We can take the partial derivative of the above objective function $\psi_{TL}(\mathrm{X},\mathrm{Y})$ with respect to X and Y as follows:

$$\frac{\partial(\psi_{TL}(\mathrm{X},\mathrm{Y}))}{\partial(X)} = 2\left(X - \hat{X} + \lambda^{com}W^{com}X + \lambda^{atc}W^{atc}X + \gamma\mathrm{X} - \gamma YY^T\right),$$

$$\frac{\partial(\psi_{TL}(\mathrm{X},\mathrm{Y}))}{\partial(Y)} = 2\left(Y - \hat{Y} + \lambda^{pro}W^{pro}Y - 2\gamma\mathrm{YX} + 2\gamma YY^T Y\right).$$

After some algebraic transformations, the updating rules can be determined to learn X and Y through gradient descent [48] as follows:

$$X(t+1) = X(t) - \theta * 2\left(X(t) - \hat{X} + \lambda^{com}W^{com}X(t) + \lambda^{atc}W^{atc}X(t) + \gamma\mathrm{X(t)} - \gamma Y(t)Y(t)^T\right),$$

$$Y(t+1) = Y(t) - \theta * 2\left(Y(t) - \hat{Y} + \lambda^{pro}W^{pro}Y(t) - 2\gamma\mathrm{Y(t)X(t)} + 2\gamma Y(t)Y(t)^T Y(t)\right),$$

where θ is the learning rate (set to 0.01 in this study). The matrices X and Y are updated according to the above rules until they converge. Finally, potential drug–target interactions are prioritized with the entities in matrix Y, and the synergistic effect between drug i and drug j is scored with $X_{ij} + X_{ji}$, because matrix X may not be symmetric. The DSML exploration process is summarized in Algorithm 1. We uploaded an R-language software package for executing the DSML method to GitHub at https://github.com/Chenxin-99/DSML that can be used to perform cross-validation and model prediction with optimal parameters to reproduce our results.

Algorithm 1 (DSML).

Input: Effective synergistic drug combination; ATC code; protein-protein interactions; drug-target protein interactions

Output: Synergy scores of drug pairs

Preprocessing phase:

1. Construct matrices $\hat{X}$ and $\hat{Y}$

2. Calculate the drug anatomical therapeutic similarity and construct matrix S^{atc}

3. Calculate the protein topological similarity and construct matrix S^{pro}

4. $S^{com} = X$

5. $W^{atc} = I^{n*n} - D^{atc^{-1/2}} S^{atc} D^{atc^{-1/2}}$, $W^{pro} = I^{m*m} - D^{pro^{-1/2}} S^{pro} D^{pro^{-1/2}}$, $W^{com} = I^{n*n} - D^{com^{-1/2}} S^{com} D^{com^{-1/2}}$

Prediction phase:

6. Solve the minimization problem

$\mathrm{X,Y} = argmin_{\mathrm{X,Y}} \psi_{TL}(X,Y)$ using gradient descent:

For matrix X:

$$X(t+1) = X(t) - \theta * 2\left(X(t) - \hat{X} + \lambda^{com} W^{com} X(t) + \lambda^{atc} W^{atc} X(t) + \gamma \mathrm{X(t)} - \gamma Y(t) Y(t)^T\right)$$

For matrix Y:

$$Y(t+1) = Y(t) - \theta * 2\left(Y(t) - \hat{Y} + \lambda^{pro} W^{pro} Y(t) - 2\gamma \mathrm{Y(t)X(t)} + 2\gamma Y(t) Y(t)^T Y(t)\right)$$

when $\|X(t+1) - X(t)\|^2 \leq threshold$

7. The synergistic score of a drug pair i and j is measured by $X_{ij}+X_{ji}$

3 Experiments

In these experiments, we first compare DSML with two other methods through cross-validation experiments. Then, we present several examples to show how the reconstruction of drug–target interactions and incorporation of knowledge from various sources can benefit the prediction of synergistic drug combinations. Finally, we provide case studies that confirm the ability of DSML to predict drug synergy.

3.1 Data Preparation and Parameter Selection

We adopted the dataset curated in our previous study [25], which contains 173 effective drug combinations, 129 protein–protein interactions and 449 drug–target interactions involving in 218 targets and 139 drugs. This benchmark database was curated from the DCDB 2.0 [31], HPRD [34], and DrugBank 5.0 [33], and it can be obtained from https://github.com/Chenxin-99/DSML. In the experiment, three cross-validation strategies, including unbalanced five-fold cross-validation (U-FFCV), balanced five-fold cross-validation (B-FFCV) and leave-one-out

cross-validation (LOOCV), were employed to evaluate the predictive results of DSML and two competing methods. These three cross-validation strategies are illustrated in detail in Supplementary Fig. 1. Supplementary Fig. 1A shows the matrix of the known drug combinations obtained from the benchmark dataset. As shown in Supplementary Fig. S1B and C, we first divide all the known effective drug combinations into five equal sized subsets; then, four of these subsets are used as the training set, while the remaining subset and all the unknown drug pairs are used as the testing set in U-FFCV [49, 50]. For B-FFCV, the remaining subset and an equal number of negative samples selected from unknown drug pairs through a simple random sampling algorithm are used as the testing set [51, 52]. As shown in Supplementary Fig. 1D, in the LOOCV experiment, we take each known combination in the testing set in turn as a positive sample, while all the unknown pairs in the testing set are taken as negative samples [43].

In the experiments, under different ranking cutoffs, the false-positive ratios (FPRs) and the true-positive ratios (TPRs) can be computed to draw a receiver-operating characteristic curve (ROC) and calculate the area under the ROC curve (AUC) [53]. Furthermore, to heavily punish highly ranked negative samples, we also utilize the area under the precision-recall curve (AUPR) to evaluate the

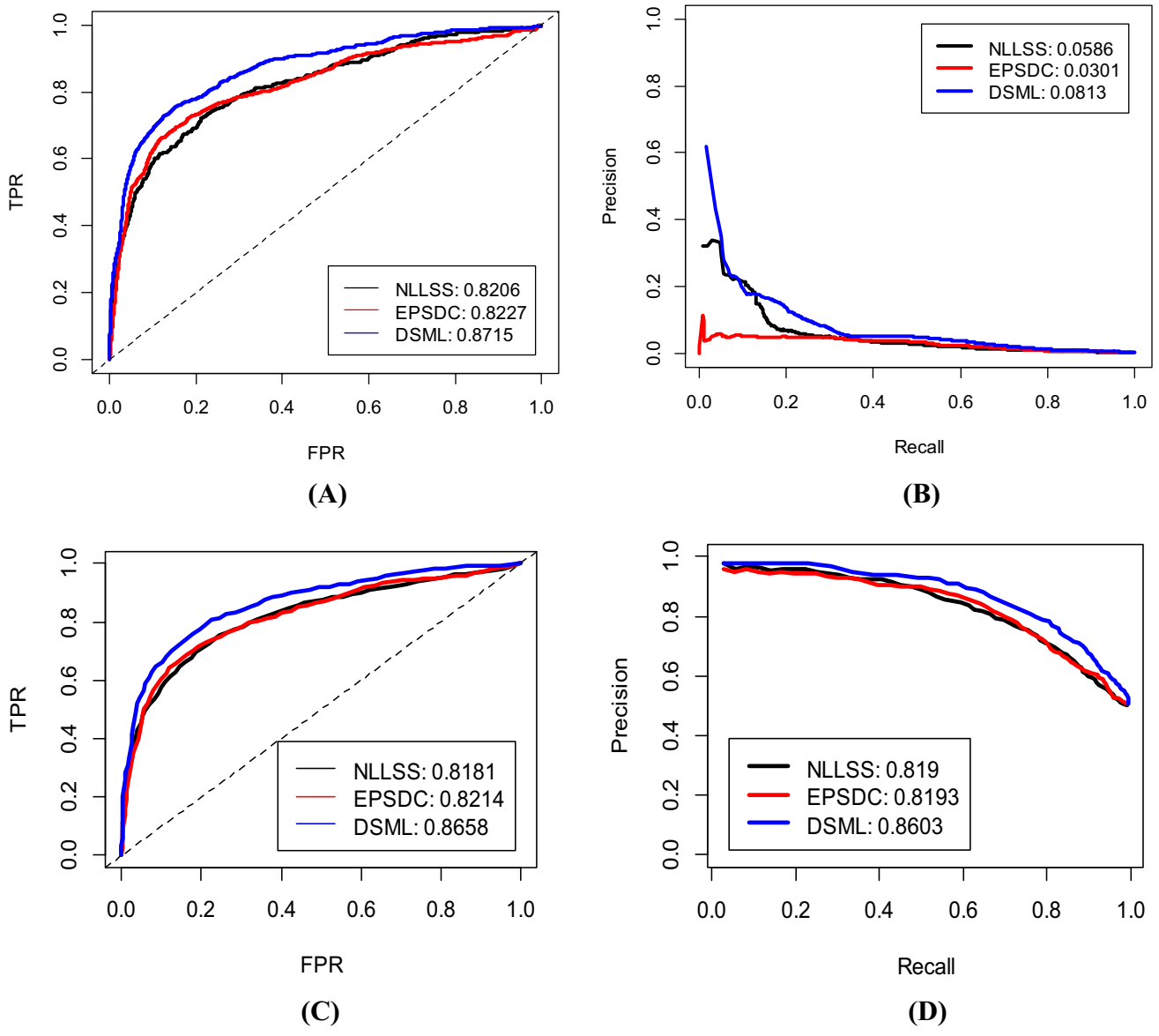

Fig. 2 Drug synergy prediction performance in the FFCV experiment. **a** ROC curves in U-FFCV. **b** PR curves in U-FFCV. **c** ROC curves in B-FFCV. **d** PR curves in B-FFCV

synergistic drug combinations prediction ability of these computational models. The precision-recall (PR) curve was plotted using the precision and recall values at different ranking cutoffs [54]. Moreover, the F1-score is a statistical index used to measure the performance of binary model and is calculated by the harmonic mean of precision and recall, defined formally by $F1-score = 2 * \frac{precision*recall}{precision+recall}$. The five-fold cross-validation runs were performed 10 times to set the DSML parameters (λ^{com}, λ^{atc}, and λ^{pro}). Since the proposed computational model is valid without restriction of $\lambda^{com} + \lambda^{atc} + \lambda^{pro} = 1$ (unlike MDAGRF [43] in which the sum of parameters must be 1), we utilized a grid search to obtain the best parameter combination from the values: $\lambda^{com}/\lambda^{atc}/\lambda^{pro} \in \{0.001, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9\}$. This strategy has been used in other studies [55, 56] to select parameter combinations from a larger space. We set λ^{com}, λ^{atc}, and λ^{pro} to 0.1, 0.1, and 0.7, respectively. Furthermore, the parameters of compared methods were set to the default values suggested in those studies (e.g., maximum length = 4 for EPSDC and $\eta_A = 0.3, \eta_P = 0.3$ for NLLSS).

3.2 Cross-validation Experiments

Three cross-validation experiments, including U-FFCV, B-FFCV and LOOCV, were performed to compare DSML to two other methods, NLLSS and EPSDC. After training the models, we adopted the AUC and AUPR to evaluate the prediction results of each method by prioritizing drug pairs in the testing set. Figure 2 shows the scoring drug synergy performance on FFCV, which included U-FFCV and B-FFCV. As shown in Fig. 2a, in the U-FFCV experiment, the AUCs of NLLSS and EPSDC are 0.8206 and 0.8227, respectively, while the AUC of the multitask learning method DSML is 0.8693; in other words, DSML outperforms the other models in terms of AUC. Furthermore, a comparison of the AUPR among the different methods in the U-FFCV experiment is shown in Fig. 2b. The multitask learning method DSML (AUPR of 0.0816) achieves the best prediction performance among all the methods, improving on NLLSS (AUPR of 0.0586) and EPSDC (AUPR of 0.0301) by a large margin because of the additional training information transferred from the predictive drug–target interactions. Comparing the PR curves and ROC curves, we can observe that the precision is quite low at a reasonably high recall and AUPR is quite low at a high AUC value in the U-FFCV experiment. A similar phenomenon has appeared in many other studies [50, 57]. The main reason for these results is imbalance in the benchmark dataset. We discuss this topic in detail in the https://github.com/Chenxin-99/DSML. As shown in Fig. 2c, d, in the B-FFCV experiment, DSML significantly outperformed the two other methods by at least 4.44% in terms of AUC and 4.1% in terms of AUPR. Moreover, the precision, recall, and F1-score comparisons among the different methods at different thresholds in the FFCV experiment are shown in Supplementary Fig. 2. As shown in Supplementary Fig. 2A–C, we found an increase in prediction performance of at least 0.476% in terms of precision, 3.55% in terms of recall, and 0.89% in terms of F1-score on U-FFCV. Similarly, we found increase in prediction performance (at least 1.68% in terms of precision, 0.353% in terms of recall, and 0.615% in terms of F1-score) on B-FFCV. In other words, DSML still outperforms the other two approaches in terms

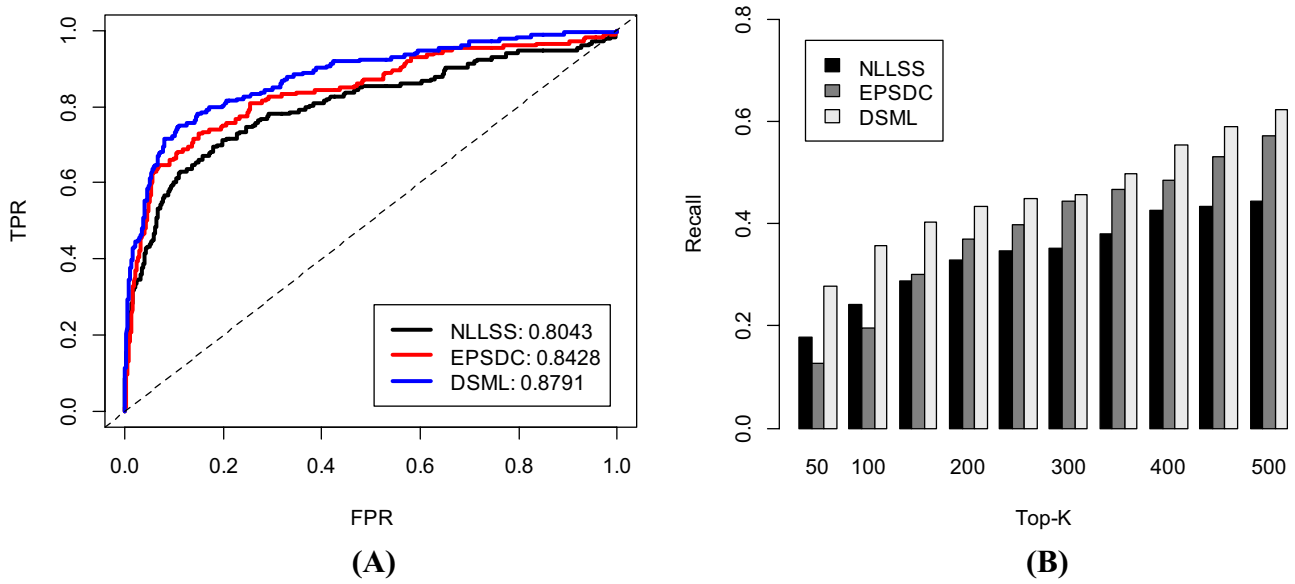

Fig. 3 Drug synergy scoring performance on the LOOCV experiment. **a** ROC curves. **b** Recalls values

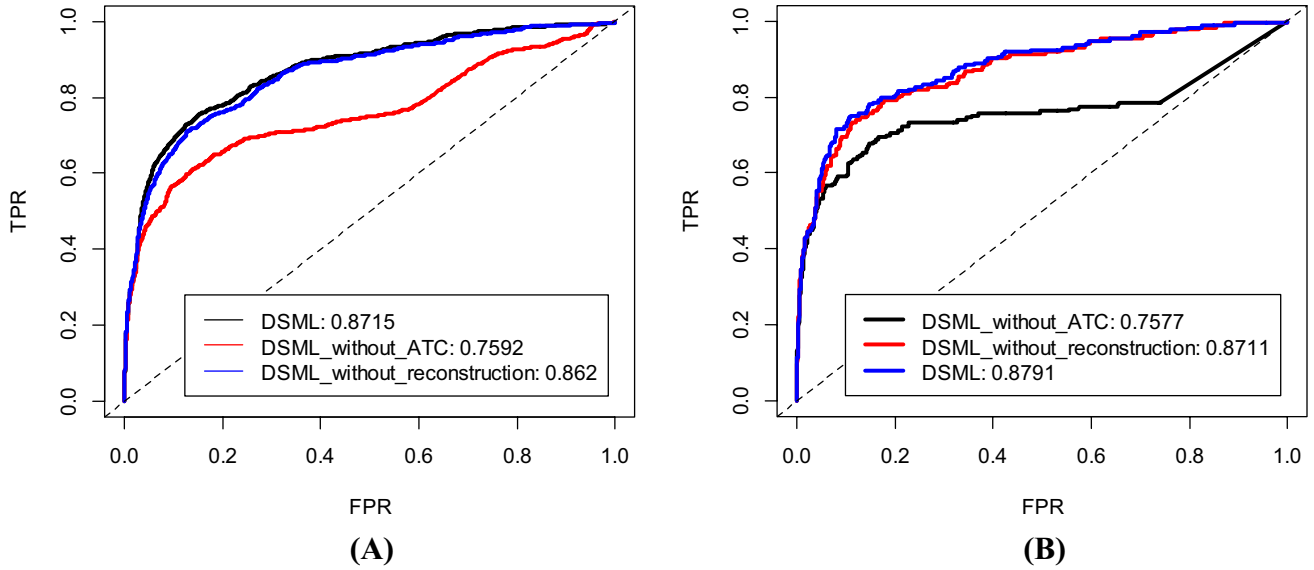

Fig. 4 Performances of three DSML variants. **a** ROC curves of the three DSML variants on U-FFCV. **b** ROC curves of the three DSML variants on LOOCV

of precision, recall, and F1-score at various thresholds on both B-FFCV and U-FFCV. The above experiments suggest that DSML has a much better prediction capacity than the other state-of-the-art methods, when predicting drug synergy.

Compared with the FFCV experiment, the LOOCV experiment produces training data that are more similar to real data. To further demonstrate that DSML can outperform other computational methods for predicting drug synergy, we evaluated the models for scoring drug synergy on LOOCV. We used only AUC and recall as metrics in LOOCV, because of the extreme imbalance between positive samples and negative samples. Figure 3a, b shows the overall results. DSML consistently outperforms the other two methods in terms of AUC and recalls at all thresholds. Figure 3a shows that DSML achieves AUC 0.8791, versus 0.8043 for NLLSS and 0.8428 for EPSDC. Similarly, as shown in Fig. 3b, DSML is superior to the other methods, with recall values from the top 50 to the top 500. The reason for these results may be that DSML effectively incorporates various sources of knowledge through graph regularization terms and dual least squares method.

3.3 Importance of Various Source Knowledge

To validate the usefulness of anatomical therapeutic groups and reconstruction of drug–target protein interactions, three DSML variants, DSML, DSML_without_ATC, and DSML_without_reconstruction, were implemented. The first DSML was trained with all knowledge, DSML_without_ATC was trained without the drug anatomical therapeutic similarity, while DSML_without_reconstruction was trained without reconstruction of drug–target protein interactions; namely, drug–target protein and protein–protein interaction knowledge was ignored in DSML_without_reconstruction. As shown in Fig. 4, the DSML variant performed better than the DSML_without_ATC and DSML_without_reconstruction variants in additional experiments on U-FFCV and LOOCV. In terms of AUC, DSML outperformed DSML_without_ATC and DSML_without_reconstruction on U-FFCV by 11.23% and 0.95% (Fig. 4a) and by 12.14% and 0.8% on LOOCV (Fig. 4b), respectively. These results may illustrate that the drug synergy prediction performance can be improved by incorporating various sources of knowledge and by reconstructing drug–targets.

3.4 Novel Synergistic Drug Combination

To further demonstrate the potential of DSML for screening synergistic drug combinations, we performed an additional experiment based on the current version of the database and then, treated the unknown combinations (neutral samples) as candidate sets for validation. Finally, we ranked the candidate drug pairs based on the synergy scores calculated by DSML. Table 2 lists the top five combinations ranked by DSML based on multitask learning. To obtain more useful information, we referred to the relevant literature. The gemcitabine and 5-fluorouracil combination ranked first and could be useful in combating advanced pancreatic cancer with an acceptable

Table 2 Combinations with the top five synergy scores calculated by DSML

Rank	Drug name	Targets	Drug name	Targets	Disease
1	Gemcitabine	Ribonucleoside-diphosphate reductase large subunit, Thymidylate synthase, UMP-CMP kinase, DNA	Fluorouracil	Thymidylate synthase, DNA, RNA	Advanced pancreatic cancer
2	Cyclophosphamide	Nuclear receptor subfamily 1 group I member 2, DNA	Oxaliplatin	DNA	Locally advanced cervical cancer
3	Cyclophosphamide	Nuclear receptor subfamily 1 group I member 2, DNA	Erlotinib	Epidermal growth factor receptor, Nuclear receptor subfamily 1 group I member 2	Metastatic breast cancer
4	Methotrexate	Dihydrofolate reductase, Thymidylate synthase, Bifunctional purine biosynthesis protein PURH	Cisplatin	DNA-3-methyladenine glycosylase, Alpha-2-macroglobulin, Serotransferrin, Copper transport protein ATOX1, DNA	Invasive bladder cancer
5	Valproic acid	Short/branched chain-specific acyl-CoA dehydrogenase, mitochondrial, Histone deacetylase 9, 2-oxoglutarate dehydrogenase, mitochondrial, Succinate-semialdehyde dehydrogenase, mitochondrial, sodium channel protein, Histone deacetylase 2, Peroxisome proliferator-activated receptor alpha, Peroxisome proliferator-activated receptor delta, Peroxisome proliferator-activated receptor gamma	Carbamazepine	Voltage-gated sodium channel alpha subunit, Neuronal acetylcholine receptor subunit alpha-4, Nuclear receptor subfamily 1 group I member 2	Rapid-cycling bipolar disorder

toxicity profile [58]. The oxaliplatin and cyclophosphamide combination, which ranked second, has synergistic action and has been proven to be effective for treating locally advanced cervical cancer [59]. Montagna et al. [60] reported that patients with metastatic breast cancer could receive metronomic oral capecitabine and cyclophosphamide plus bevacizumab and erlotinib. A previous study reported that the use of neoadjuvant cisplatin, methotrexate, and vinblastine (CMV) chemotherapy could improve invasive bladder cancer treatment outcomes. The carbamazepine and valproic acid combination, which ranked fifth, may show synergistic anticonvulsant efficacy [61]. Moreover, we provide a list of the top 100 potential drug combinations at https://github.com/Chenxin-99/DSML.

4 Discussion

In this study, in contrast to other approaches, we explored the use of multitask learning to improve synergistic drug combination predictions by effectively reconstructing drug–target protein interactions. The ablation experiment between DSML and DSML_without_reconstruction, which is a single-task learning method, demonstrated that multitask can enrich the training information of synergistic drug combinations by predicting drug–target interactions, resulting in an overall improvement. The comparisons with the EPSDC and NLLSS methods indicated that although the other two methods can incorporate various types of information, DSML seems to be a better computational approach for calculating a synergy score. The main factors that contribute to the success of DSML may be that it differs significantly from the previous methods. First, DSML effectively considers the characteristics of synergistic agents in various types of information, including drug–target protein interactions and ATC codes. Second, to predict potential candidate combinations, DSML can use both the known and experimental synergistic drug combinations as seed datasets. In addition, many studies have shown that synergistic drug combinations can be screened out without negative samples; DSML performs this as well. Moreover, we not only integrated knowledge from various sources through the establishment of a computational model but also reconstructed drug–target protein interactions to further improve the performance of screening synergistic drug combinations. As stated above, we employed cross-validation experiments on the benchmark datasets. As a result, throughout all the verification schemes, DSML was proved to be reliable and showed better prediction performances, which may indicate that reconstructing drug–target protein interactions contributes to screening synergistic drug combinations.

However, some limitations of DSML should be mentioned. First, although this method includes the reconstruction of drug–target protein interactions to improve the prediction performance, it is difficult to obtain excellent computational performance, since the set of protein–protein interaction is sparse and noisy. In future work, we plan to incorporate more drug features to improve the predictive performance of synergistic drug combinations. For example, miRNAs with proven biological functions have become increasingly popular drug targets [62]. Drug–disease association is also helpful for understanding the mechanisms of drug therapy in disease [63]. Therefore, incorporating the accumulated miRNA–drug interactions and disease–drug associations may further improve drug synergy prediction. Second, DSML can currently screen only pair-wise synergistic drug combinations, not drug combinations of more than two drugs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12539-021-00422-x.

Acknowledgements This work has been supported by the National Natural Science Foundation of China (Grant Nos. 62002154, 61873089, and 61502221), Hunan Provincial Natural Science Foundation of China (Grant No. 2019JJ50520), Research Foundation of Hunan Educational Committee (Grant No. 20C1579), and Scientific Research Startup Foundation of University of South China (Grant No. 190XQD096).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

1. Spiro Z, Kovacs IA, Csermely P (2008) Drug-therapy networks and the prediction of novel drug targets. J Biol 7(6):1–5. https://doi.org/10.1186/jbiol81
2. Chou TC (2010) Drug combination studies and their synergy quantification using the Chou-Talalay method. Can Res 70(2):440–446. https://doi.org/10.1158/0008-5472.CAN-09-1947
3. Lecca P, Priami C (2013) Biological network inference for drug discovery. Drug Discov Today 18(5):256–264. https://doi.org/10.1016/j.drudis.2012.11.001
4. Lötsch J, Geisslinger G (2011) Low-dose drug combinations along molecular pathways could maximize therapeutic effectiveness while minimizing collateral adverse effects. Drug Discov Today 16(23):1001–1006. https://doi.org/10.1016/j.drudis.2011.10.003
5. Meyer CT, Wooten DJ, Paudel BB, Bauer J, Hardeman KN, Westover D, Lovly CM, Harris LA, Tyson DR, Quaranta V (2019) Quantifying drug combination synergy along potency and efficacy axes. Cell Syst 8(2): 97–108. e116. https://doi.org/https://doi.org/10.1016/j.cels.2019.01.003
6. Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, Grange F, Mortier L (2015) Improved overall survival in melanoma with combined dabrafenib and trametinib. N Engl J Med 372(1):30–39. https://doi.org/10.1056/NEJMoa1412690
7. Jia J, Zhu F, Ma X, Cao ZW, Li YX, Chen YZ (2009) Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov 8(2):111–128. https://doi.org/10.1038/nrd2683
8. Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L (2020) DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Res 48(D1):D871–D881. https://doi.org/10.1093/nar/gkz1007
9. Shekhar C (2008) In silico pharmacology: computer-aided methods could transform drug development. Chem Biol 15(5):413–414. https://doi.org/10.1016/j.chembiol.2008.05.001
10. Yu Y, Li M, Liu L, Li Y, Wang J (2019) Clinical big data and deep learning: applications, challenges, and future outlooks. Big Data Mining Anal 2(4): 288–305. https://doi.org/https://doi.org/10.26599/BDMA.2019.9020007
11. Zhao XM, Iskar M, Zeller G, Kuhn M, Van NV, Bork P (2011) Prediction of drug combinations by integrating molecular and pharmacological data. PLoS Comput Biol 7(12):e1002323. https://doi.org/10.1371/journal.pcbi.1002323
12. Zou J, Ji P, Zhao YL, Li LL, Wei YQ, Chen YZ, Yang SY (2012) Neighbor communities in drug combination networks characterize synergistic effect. Mol BioSyst 8(12):3185–3196. https://doi.org/10.1039/C2MB25267H
13. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: A new perspective. Neurocomputing 300:70–79. https://doi.org/10.1016/j.neucom.2017.11.077
14. Liu H, Zhang W, Nie L, Ding X, Zou L (2019) Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network. BMC Bioinformatics 20(1):1–12. https://doi.org/10.1186/s12859-019-3288-1
15. Sheng Z, Sun Y, Yin Z, Tang K, Cao Z (2017) Advances in computational approaches in identifying synergistic drug combinations. Brief Bioinform 19(6):1172–1182. https://doi.org/10.1093/bib/bbx047
16. Li S, Zhang B, Zhang N (2011) Network target for screening synergistic drug combinations with application to traditional Chinese medicine. BMC Syst Biol 5(S1):S10. https://doi.org/10.1186/1752-0509-5-S1-S10
17. Lee JH, Kim DG, Bae TJ, Rho K, Kim JT, Lee JJ, Jang Y, Kim BC, Park KM, Kim S (2012) CDA: combinatorial drug discovery using transcriptional response modules. PLoS ONE 7(8):e42573. https://doi.org/10.1371/journal.pone.0042573
18. Yang J, Tang H, Li Y, Zhong R, Wang T, Wong S, Xiao G, Xie Y (2015) DIGRE: drug-induced genomic residual effect model for successful prediction of multidrug effects. CPT Pharmacometr Syst Pharmacol 4(2):91–97. https://doi.org/10.1002/psp4.1
19. Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, Xiao G, Li Y, Allen J, Zhong R (2014) A community computational challenge to predict the activity of pairs of compounds. Nat Biotechnol 32(12):1213–1222. https://doi.org/10.1038/nbt.3052
20. Xu KJ, Song J, Zhao XM (2012) The drug cocktail network. BMC Syst Biol 6(1):S5. https://doi.org/10.1186/1752-0509-6-S1-S5
21. Wang YY, Xu KJ, Song J, Zhao XM (2012) Exploring drug combinations in genetic interaction network. BMC Bioinformatics 13(7):S7. https://doi.org/10.1186/1471-2105-13-S7-S7
22. Xu Q, Xiong Y, Dai H, Kumari KM, Xu Q, Ou HY, Wei DQ (2017) PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm. J Theor Biol 417:1–7. https://doi.org/10.1016/j.jtbi.2017.01.019
23. Sun Y, Sheng Z, Ma C, Tang K, Zhu R, Wu Z, Shen R, Feng J, Wu D, Huang D (2015) Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. Nat Commun 6(1):1–10. https://doi.org/10.1038/ncomms9481

24. Chen X, Ren B, Chen M, Wang Q, Zhang L, Yan G (2016) NLLSS: predicting synergistic drug combinations based on semi-supervised learning. PLoS Comput Biol 12(7):e1004975. https://doi.org/10.1371/journal.pcbi.1004975
25. Ding P, Yin R, Luo J, Kwoh CK (2019) Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological and network knowledge. IEEE J Biomed Health Inform 23(3):1336–1345. https://doi.org/10.1109/JBHI.2018.2852274
26. Ding P, Ouyang W, Luo J, Kwoh CK (2020) Heterogeneous information network and its application to human health and disease. Brief Bioinform 21(4):1327–1346. https://doi.org/10.1093/bib/bbz091
27. Zhang X, Song J, Bork P, Zhao X (2016) The exploration of network motifs as potential drug targets from post-translational regulatory networks. Sci Rep 6(1):20558–20558. https://doi.org/10.1038/srep20558
28. Wang Y, Nacher JC, Zhao X (2012) Predicting drug targets based on protein domains. Mol BioSyst 8(5):1528–1534. https://doi.org/10.1039/c2mb05450g
29. Zeng X, Zhu S, Hou Y, Zhang P, Li L, Li J, Huang LF, Lewis SJ, Nussinov R, Cheng F (2020) Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. Bioinformatics 36(9):2805–2812. https://doi.org/10.1093/bioinformatics/btaa010
30. Zeng X, Zhu S, Lu W, Liu Z, Huang J, Zhou Y, Fang J, Huang Y, Guo H, Li L (2020) Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci 11:1775–1797. https://doi.org/10.1039/c9sc04336e
31. Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X (2014) DCDB 2.0: a major update of the drug combination database. Database. https://doi.org/10.1093/database/bau124
32. Cheng F, Li W, Wu Z, Wang X, Zhang C, Li J, Liu G, Tang Y (2013) Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space. J Chem Inf Model 53(4):753–762. https://doi.org/10.1021/ci400010x
33. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z (2017) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46(D1):D1074–D1082. https://doi.org/10.1093/nar/gkx1037
34. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A (2008) Human protein reference database—2009 update. Nucleic Acids Res 37(S1):D767–D772. https://doi.org/10.1093/nar/gkn892
35. Asur S, Ucar D, Parthasarathy S (2007) An ensemble framework for clustering protein–protein interaction networks. Bioinformatics 23(13):i29–i40. https://doi.org/10.1093/bioinformatics/btm212
36. Cao B, Luo J, Liang C, Wang S, Ding P (2016) Pce-fr: a novel method for identifying overlapping protein complexes in weighted protein-protein interaction networks using pseudo-clique extension based on fuzzy relation. IEEE Trans Nanobiosci 15(7):728–738. https://doi.org/10.1109/TNB.2016.2611683
37. Lei C, Ruan J (2013) A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. Bioinformatics 29(3):355–364. https://doi.org/10.1093/bioinformatics/bts688
38. Li A, Horvath S (2007) Network neighborhood analysis with the multi-node topological overlap measure. Bioinformatics 23(2):222–231. https://doi.org/10.1093/bioinformatics/btl581
39. Lei C, Tamim S, Bishop AJ, Ruan J (2013) Fully automated protein complex prediction based on topological similarity and community structure. Proteome Sci 11(1):1–8. https://doi.org/10.1186/1477-5956-11-S1-S9
40. Zhao Y, Chen X, Yin J (2018) A novel computational method for the identification of potential miRNA-disease association based on symmetric non-negative matrix factorization and Kronecker regularized least square. Front Genet 9:324. https://doi.org/10.3389/fgene.2018.00324
41. Wang F, Huang ZA, Chen X, Zhu Z, Wen Z, Zhao J, Yan GY (2017) LRLSHMDA: Laplacian regularized least squares for human microbe–disease association prediction. Sci Rep 7(1):1–11. https://doi.org/10.1038/s41598-017-08127-2
42. Gui J, Huang DS, You Z (2008) An improvement on learning with local and global consistency. Int Conf Pattern Recogn 19:1–4. https://doi.org/10.1109/ICPR.2008.4761295
43. Luo J, Ding P, Liang C, Chen X (2018) Semi-supervised prediction of human miRNA-disease association based on graph regularization framework in heterogeneous networks. Neurocomputing 294:29–38. https://doi.org/10.1016/j.neucom.2018.03.003
44. Long M, Wang J, Ding G, Shen D, Yang Q (2014) Transfer learning with graph co-regularization. IEEE Trans Knowl Data Eng 26(7):1805–1818. https://doi.org/10.1109/TKDE.2013.97
45. Ding P, Shen C, Lai Z, Liang C, Li G, Luo J (2020) Incorporating multisource knowledge to predict drug synergy based on graph co-regularization. J Chem Inf Model 60(1):37–46. https://doi.org/10.1021/acs.jcim.9b00793
46. Petegrosso R, Park S, Hwang TH, Kuang R (2016) Transfer learning across ontologies for phenome–genome association prediction. Bioinformatics 33(4):529–536. https://doi.org/10.1093/bioinformatics/btw649
47. Ezzat A, Zhao P, Wu M, Li XL, Kwoh CK (2017) Drug-target interaction prediction with graph regularized matrix factorization. IEEE/ACM Trans Comput Biol Bioinf 14(3):646–656. https://doi.org/10.1109/TCBB.2016.2530062
48. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010. 19: 177–186. https://doi.org/https://doi.org/10.1007/978-3-7908-2604-3_16
49. Shen C, Luo J, Ouyang W, Ding P, Chen X (2020) IDDkin: Network-based influence deep diffusion model for enhancing prediction of kinase inhibitors. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa1058
50. Lin X, Quan Z, Wang ZJ, Ma T, Zeng X (2020) KGNN: knowledge graph neural network for drug-drug interaction prediction. Int Joint Conf Artif Intell 29:2739–2745
51. Chen H, Cheng F, Li J (2020) iDrug: integration of drug repositioning and drug-target prediction via cross-network embedding. PLoS Comput Biol 16(7):e1008040. https://doi.org/10.1371/journal.pcbi.1008040
52. Zeng X, Zhu S, Liu X, Zhou Y, Cheng F (2019) deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics 35(24):5191–5198. https://doi.org/10.1093/bioinformatics/btz418
53. Muschelli J (2020) ROC and AUC with a binary predictor: a potentially misleading metric. J Classif 37(3):696–708. https://doi.org/10.1007/s00357-019-09345-1
54. Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. Int Conf Mach Learn 23:233–240. https://doi.org/10.1145/1143844.1143874
55. Fan J, Cheng J (2018) Matrix completion by deep matrix factorization. Neural Netw 98:34–41. https://doi.org/10.1016/j.neunet.2017.10.007
56. Chen H, Li J (2019) Modeling Relational Drug-target-disease interactions via tensor factorization with multiple web sources. World Wide Web Conf 19:218–227. https://doi.org/10.1145/3308558.3313476
57. Wan F, Hong L, Xiao A, Jiang T, Zeng J (2019) NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. Bioinformatics 35(1):104–111. https://doi.org/10.1093/bioinformatics/bty543

58. Gennatas C, Michalaki V, Mouratidou D, Tsavaris N, Andreadis C, Photopoulos A, Voros D (2006) Gemcitabine combined with 5-fluorouracil for the treatment of advanced carcinoma of the pancreas. In vivo 20(2):301–305. https://doi.org/10.1089/hum.2006.17.362
59. Gutierrez-Delgado F, Lopez-Mariscal A, Maldonado-Hernandez H, Luna-Benitez I, Salazar-Macias F, Aceves-Escarcega A, Delgadillo-Hernandez (2005) Oxaliplatin and cyclophosphamide as neoadjuvant chemotherapy (NACT) followed by surgery for patients with locally advanced cervical cancer (LACC). A preliminary report. J Clin Oncol 23(16_suppl): 5173–5173. https://doi.org/https://doi.org/10.1200/jco.2005.23.16_suppl.5173
60. Montagna E, Cancello G, Bagnardi V, Pastrello D, Dellapasqua S, Perri G, Viale G, Veronesi P, Luini A, Intra M (2012) Metronomic chemotherapy combined with bevacizumab and erlotinib in patients with metastatic HER2-negative breast cancer: clinical and biological activity. Clin Breast Cancer 12(3):207–214. https://doi.org/10.1016/j.clbc.2012.03.008
61. Ketter TA, Pazzaglia PJ, Post ARM (1992) Synergy of carbamazepine and valproic acid in affective illness: case report. J Clin Psychopharmacol 12(4):276–281. https://doi.org/10.1097/00004714-199208000-00011
62. Chen X, Xie W, Xiao P, Zhao X, Yan H (2017) mTD: a database of microRNAs affecting therapeutic effects of drugs. J Genet Genomics 44(5):269–271. https://doi.org/10.1016/j.jgg.2017.04.003
63. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers T, CMattingly C. (2016) The comparative toxicogenomics database: update 2017. Nucleic Acids Res 45(D1):D972–D978. https://doi.org/10.1093/nar/gkw838