OXFORD

Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects

Kunjie Fan 🆔, Lijun Cheng and Lang Li

Corresponding author. Lang Li, Department of Biomedical Informatics, College of Medicine, The Ohio State University, 250B Lincoln Tower, 1800 Cannon Drive, Columbus, OH 43210, USA. Tel.: +1-614-292-4778; E-mail: Lang.Li@osumc.edu

Abstract

Drug combinations have exhibited promising therapeutic effects in treating cancer patients with less toxicity and adverse side effects. However, it is infeasible to experimentally screen the enormous search space of all possible drug combinations. Therefore, developing computational models to efficiently and accurately identify potential anti-cancer synergistic drug combinations has attracted a lot of attention from the scientific community. Hypothesis-driven explicit mathematical methods or network pharmacology models have been popular in the last decade and have been comprehensively reviewed in previous surveys. With the surge of artificial intelligence and greater availability of large-scale datasets, machine learning especially deep learning methods are gaining popularity in the field of computational models for anti-cancer drug synergy prediction. Machine learning-based methods can be derived without strong assumptions about underlying mechanisms and have achieved state-of-the-art prediction performances, promoting much greater growth of the field. Here, we present a structured overview of available large-scale databases and machine learning especially deep learning methods in computational predictive models for anti-cancer drug synergy prediction. We provide a unified framework for machine learning models and detail existing model architectures as well as their contributions and limitations, shedding light into the future design of computational models. Besides, unbiased experiments are conducted to provide in-depth comparisons between reviewed papers in terms of their prediction performance.

Key words: drug synergy; cancer; machine learning; deep learning; pharmacogenomics

Introduction

The superior efficacy of combination compared to monotherapies in patients with cancer has led to their greater popularity [1]. Drug combinations can potentially be used in lower doses and reduce toxicity and adverse side effects [2]. For example, targeting MEK and BRAF as two components of mitogen-activated protein kinase pathway in patients with melanomas harboring BRAF V600E mutations leads to less toxic effect and resistance compared to single-drug therapy [3, 4]. Currently, most effective anti-cancer drug combinations are developed through intensive trial and error without consideration of mechanisms of action, and the experimental screening of synergistic drug combinations requires prohibitively enormous search space. The development of reliable computational models for predicting anti-cancer drug synergy is therefore desirable. Currently, there are no existing published reviews on summarizing anti-cancer drug synergy prediction methods. Therefore, we aim to provide a comprehensive review focusing on the discussion of anti-cancer drug synergy prediction methods that will help both cancer biologists and computational biologists.

Kunjie Fan is a PhD candidate at the Department of Biomedical Informatics of The Ohio State University. His study focuses on the development of computational models for predicting anti-cancer drug synergy.

Lijun Cheng is an assistant professor at the Department of Biomedical Informatics of The Ohio State University. Her group's interest is to develop system pharmacology models to predict drug response for either single drug or drug combinations.

Lang Li is a professor at the Department of Biomedical Informatics of The Ohio State University. His group's interest includes literature-based text mining, translational drug interaction study and system pharmacology.

Predicting drug synergy in general (not specific to cancers) has been studied for a long time and shares similarities with anti-cancer drug synergy prediction. Several reviews have introduced recent advances in computational approaches to the identification of synergistic drug combinations in general. Given huge amounts of available literature, it is necessary to briefly introduce some core methods used in general drug synergy prediction tasks, as computational models for predicting cancer-specific drug synergy typically stem from those general methods with special considerations on cancer-related characteristics. In one recent review, Li *et al.* primarily summarized biomolecular network-based models and categorized them into unsupervised, semi-supervised and supervised groups according to whether labeled known drug combinations were used to guide their construction [5]. Ryall and Tan described system biology-based predictive methods, especially those modeling signaling networks [6], and Bulusu's team reviewed a subset of computational methods that included mathematical and gene expression- and pathway-based approaches [7]. Tonekaboni's research group provided a comprehensive study of computational approaches for combination therapies [8], discussing quantification methods and methodologies for optimal designs of combinatorial treatment assays in addition to three categories of predictive models, including mathematical and machine learning methods as well as stochastic search algorithms.

Those four reviews mainly summarized hypothesis-driven explicit mathematical methods or network pharmacology models, popular in the last decade and notable for their involvement of signaling networks and perturbation data to elucidate underlying mechanisms in a network perspective. In general, these models simultaneously modulate multiple proteins in a network to replicate the effects of a drug action [9]. Differential equation models are utilized to derive network dynamics from molecular profiles of perturbed cellular systems [10, 11], aiming to predict quantitative outcomes of combinatorial perturbations. However, the requirement for perturbation experiments to construct these two models limits the possibility of their large-scale application. Avoiding perturbation profiles, Flobak's group proposed an algorithm to build a logical model based on known signaling pathways and baseline AGS gastric cancer cell line profiles that could be used to simulate the effects of individual drugs and drug combinations for the identification of potential synergistic drug combinations [12].

Instead of building signaling networks from perturbation data to make predictions, some investigators have sought to predict drug synergy directly from the network topology or sub-networks affected by drug targets. Based on the assumption that the effects of drug combinations should depend on the interaction of their targets in a network, Yin *et al.* modeled the effects of drug combinations along with their targets interacting in a network to elucidate the relationships between the network topology and drug combination effects [13]. Molecular interaction networks from protein–protein interactions (PPIs) and protein–DNA interactions [14], from gene modules [15] and from drug–target interactions (DTIs) [16] are also constructed to be used for predicting synergistic drug combinations.

These hypothesis-driven explicit mathematical methods or network pharmacology models have been popular in the last decade and have greatly advanced the development of predictive approaches for combination therapies, both for general drug combinations and anti-cancer drug combinations. However, several factors may limit their utility, including the small size and accuracy of the datasets used in their construction, the possible choice of an errant hypothesis to guide construction, their

inability to model multiple cancer types to assess drug synergy and the absence of standard guidance to process transcriptional expression data. The construction of explicit mathematical or network pharmacology modes often starts with a reasonable hypothesis of drug synergy that could, in fact, be completely incorrect. Thus, a hypothesis-free model might be preferable, especially when modeling large-scale complex datasets [17]. Furthermore, the typical optimization of these models using a single cancer type and their inability to model multiple types simultaneously limits their utility in studying relationships among different cancer types in terms of drug synergy. Finally, these models commonly use drug-treated transcriptional expression data for such information as the mechanism of drug action underlying a biological signaling pathway. However, the absence of a standard rule to process the transcriptional expression data leads to the selection of various significant differential expression genes among individual researchers and the construction of different networks and difficulty in final interpretation [5]. These limitations of explicit mathematical and network pharmacology models suggest the need for better computational models to identify anti-cancer synergistic anti-cancer drug combinations.

In recent years, the greater availability of large-scale datasets and exponential growth in computing power have led to the increased popularity of machine learning methods in many areas of computational biology and bioinformatics. Machine learning involves the study of a set of algorithms designed to allow automatic discernment, or learning, of knowledge from data that can then be accurately generalized to new, unseen data [18]. Machine learning methods are a great choice in computational biology because they can derive predictive models without requiring strong assumptions about underlying mechanisms, and they have proven successful in numerous fields, including image classification [19], speech recognition [20] and protein function prediction [21]. Deep learning methods, a subfield of machine learning, also benefit from the increasing number of available public datasets. Their ability to learn abstract representations from high-dimensional data is useful in solving complex tasks, and they are widely used to improve predictive performance. These techniques have enabled unprecedented breakthroughs in many areas, including image processing, speech recognition and text classification, performing as well as or better than human experts [22].

The increasing publication of large-scale anti-cancer drug combination screening datasets enables the development of reliable machine learning or deep learning methods for the prediction of anti-cancer drug synergy and promotes much greater growth of the field. In this review, we focus primarily on machine learning-based models, especially supervised models, which utilize labeled known drug combinations to guide the model's learning process. We categorize the models as either traditional machine learning methods or deep learning methods, discussing their architecture and features as described in each publication and detailing their contributions and possible limitations. We also conducted unbiased experiments to compare the prediction performance between reviewed methods, providing in-depth performance analysis to guide the future design of predictive models.

Large-scale public anti-cancer drug combination datasets

The number of large-scale anti-cancer drug combination datasets that promote the development of machine learning

Table 1. Existing large-scale anti-cancer drug combination screening datasets

Source	Number of cell lines	Number of tissues	Number of drugs	Number of combinations	Number of data	Experimental design	Synergy metric
AstraZeneca-Sanger DREAM	85	6	118	910	11 576	5×5	Loewe score
Merck & Co.	39	6	38	583	23 062	4×4	Loewe score
NCI-ALMANAC	60	10	104	5232	304 549	3×3 or 5×3	ComboScore

The first four columns record the unique number of cell lines, tissues, single drugs and drug combinations included in the dataset. 'Number of data' records the sample size of each dataset. 'Experimental design' indicates the dosing regimen used by each dataset in the experiments. 'Synergy metric' indicates how each dataset defines synergism. NCI, National Cancer Institute.

models has grown greatly in recent years because of the rapid development of high-throughput screening (HTS). Table 1 and the following paragraphs detail the three largest. In 2015, AstraZeneca, the well-known pharmaceutical enterprise, partnered with several institutes to launch the AstraZeneca-Sanger Drug Combination Prediction Challenge in DREAM (dialog for reverse engineering assessments and methods) community to assist the evaluation of computational strategies for the prediction of anti-cancer synergistic drug pairs and biomarkers [23]. They released 11 576 experiments from 910 combinations involving 118 drugs and 85 cell lines, establishing the largest drug combination dataset at that time and enabling the development of dozens of machine learning-based prediction models. In 2016, Merck & Co. published a large-scale oncology screen comprising 23 062 samples using a 4-by-4 dosing regimen [2]. This dataset involves 583 combinations in 39 diverse cancer cell lines and 38 unique drugs. They also performed separate single-agent screenings using eight concentrations with six replicates, which allowed definition of the edges of the combination surface by interpolation and led to a 5-by-5 concentration matrix for each sample. In 2017, the National Cancer Institute (NCI) of the United States introduced ALMANAC (a large matrix of anti-neoplastic agent combinations), the largest publicly available cancer drug combination dataset, containing synergy measurements of pair-wise combinations of 104 Food and Drug Administration (FDA)-approved drugs in 60 cancer cell lines from the NCI-60 panel [24]. For inclusion in the database, they first screened the drugs in single doses in all 60 cell lines to efficiently identify compounds with anti-proliferative activity and then screened only the drugs with above-threshold effects to obtain full dose–response matrices. In total, NCI-ALMANAC includes data of 304 549 samples covering 5232 drug combinations in 10 tissues, aggregating synergy data from three screening centers under two different experimental protocols (3-by-3 and 5-by-3). They measured the synergy state using ComboScore, a modification of Bliss independence, in which more positive ComboScore values correspond with more synergistic drug pairs.

A comprehensive database is urgently needed that collects and integrates the many small datasets and extensive HTS data that have become increasingly available. Such a database will benefit both the experimental screening and construction of computational models of anti-cancer drug combinations. DrugComb is an integrative portal of anti-cancer drug combination data in which the results of drug combination screening studies are accumulated, standardized and harmonized [25]. DrugComb has collected 437 923 combinations across 93 cancer cell lines from dozens of datasets and continues to increase the sample size by way of crowdsourcing. In the web server, four commonly used reference models for measuring synergy—Bliss

independence, highest single agent (HSA), Loewe additivity and zero interaction potency (ZIP)—are calculated and displayed [26–28]. They also provide various visualization tools for data analysis. DrugCombDB, another comprehensive database of anti-cancer drug combinations, aims to integrate various data sources, including HTS assays, manual curations from the literature, FDA Orange Book entries and failed drug combinations [29]. This database has the largest number of drug combinations to date, including 448 555 drug combinations covering 2887 unique drugs and 124 human cancer cell lines. These comprehensive integrated databases provide valuable resources for the construction of highly reliable machine learning models and facilitate the discovery of novel synergistic drug combinations for cancer therapy.

Machine learning basics

Basically, a machine learning algorithm can learn from data it is given, discerning a relationship between the features of input data and output targets. Most machine learning algorithms involve some combination of four components—a dataset, cost function, optimization procedure and model [22]. The relative independent functioning of each component in the machine learning recipe allows for their interchange and therefore a wide variety of algorithms. The cost function typically includes at least one term that causes the learning process to perform statistical estimation and may also include regularization terms to improve the generalization of the model. The most common cost function, negative log-likelihood, allows for the estimation of maximum likelihood. Gradient descent is a commonly used algorithm for optimization, especially when the model is not in linear form, such as is the case with deep-learning models.

In a machine learning paradigm, feature selection is essential to guarantee successful model building. Because the goal is to learn meaningful relationships between input features and output targets, the selected features should be intrinsically related to the target so that the learned model can make accurate novel predictions about the unseen data. For the task of predicting synergy among anti-cancer drug combinations in cell lines, input normally consists of both cell line and drug features that are predictive of drug synergy. Gene expression, mutation and copy number variation (CNV) are the three most popular cell line features, and methylation, gene essentiality and monotherapy or pathway information in the model have also been considered to improve prediction. Drug features can be roughly divided into three categories, structural features (various types of fingerprints), drug-target information and other properties, such as physicochemical properties, side effects, MACCS keys and binary toxicophore indicators. The machine learning framework represents the drug features in two common ways, either by

linking of the raw features of two candidate drugs as input into the model or by predetermining pair-wise similarities between two drugs regarding various types of features and using these processed features as input.

Output target can be a continuous synergy score or a binarized label that indicates whether there is synergy. Existing studies differ mainly in their choices of machine learning models and drug and cell line features. Some models consider drug synergy prediction as a regression problem, directly predicting raw synergy scores, and others regard it as a classification task, first binarizing synergy scores based on pre-defined thresholds.

Machine learning methods

Random forest

In this section, we discuss existing works that utilize traditional machine learning methods (non-deep learning methods). The detailed summary of all existing methods is described in Table 2. In 2017, Li's research team proposed a random forest model to predict synergistic anti-cancer drug combinations based on drug-target networks and drug-induced gene expression profiles [30]. They manually designed 18 features that included similarities among drug chemical structures, drug-target network and drug pharmacogenomics. They utilized the random forest algorithm to select best combination of features that were used to construct a prediction model and identified 28 potentially synergistic anti-cancer drug combinations. Their use of drug-induced gene expression profiles to identify differentially expressed genes for prediction was particularly noteworthy. Profiles of gene expression following treatment with different doses of drugs can reflect the biological response to drug treatment to provide further evidence of the mechanisms of drug action used for predicting drug synergy.

However, the high cost of large-scale datasets can prevent the use of information regarding drug-induced gene expression. Therefore, Li's group proposed a novel network propagation method based on gene–gene network and drug-target information to simulate post-treatment molecular features [31]. They integrated prior knowledge of drug pharmacokinetics (drug targets) and biological knowledge of gene–gene interactions into the baseline molecular profiles (gene expression, mutation, CNV, methylation) to simulate post-treatment features and generate informative features. More specifically, they utilized a network propagation method to modify baseline molecular profiles, where drug-target genes became zero and non-target genes were affected proportionally to the probabilities of their connections to the target genes for expression, methylation and CNV profiles. To improve prediction performance further, in addition to these modified molecular features, they considered monotherapy information as the input feature and utilized random forest as the classifier. This method achieved the best performance in all sub-challenges among the 160 competing teams in the AstraZeneca-Sanger Drug Combination Prediction DREAM Challenge [23]. They also showed that molecular and monotherapy features are complementary in predicting drug synergy, whereas monotherapy shows more importance when making predictions.

EXtreme Gradient Boosting

When dealing with intrinsically complex problems like the prediction of anti-cancer drug synergy, ensemble methods are popular because of their non-linear characteristics and good generalization ability. EXtreme Gradient Boosting (XGBoost), one of the most powerful ensemble methods, has been widely used to solve complex biological problems, including the prediction of drug synergy. Janizek's team introduced TreeCombo, an XGBoost-based approach to predict the synergy of novel drug combinations [32]. They included gene expression values as cell line features and drug fingerprints as well as binary toxicophore status as drug features and demonstrated that 83 of the 100 features exhibiting the highest importance values with respect to the prediction of synergy were drug-based features, thus implying the greater importance of drug than cell line features in the prediction.

Celebi *et al.* also proposed an XGBoost-based method to predict synergistic anti-cancer drug combinations and included much richer features to improve prediction accuracy and interpretability [33]. They considered multi-omics data to enrich the feature space, including gene expression, mutation and CNV. To reduce the extremely large number of gene expression features, they employed weighted gene co-expression network analysis to identify systems-level gene modules and used mean expression values of the 53 identified modules as features. They also filtered out mutations not present in the Kyoto Encyclopedia of Genes and Genomes cancer pathways and CNVs not significantly correlated with expression values across all cell lines. For drug features, they considered drug-target information, targeted pathways and drug fingerprints. They also included monotherapy information to provide additional knowledge for drug synergy prediction. Analysis of the predictions of this XGBoost-based model revealed such key regulators of tumorigenesis as TNFA and BRAF to be frequent targets in synergistic interactions and MYC to be frequently duplicated.

Sidorov *et al.* utilized XGBoost totally differently, training one unique model for each cell line (cell line specific) rather than a single model for all cell lines together [34]. Their reasoning assumed the presence of large inter-center batch effects in NCI-ALMANAC datasets. In this cell line-specific setting, no cell line features were required to distinguish among different cell lines. For drug features, they considered comprehensive information, including fingerprints, MACCS keys, ISIDA/SIRMS fragments and physicochemical properties, and they estimated reliability to identify combinations that were reliable as well as synergistic to further narrow their candidates. They found that reliability reduction permitted up to 50% reduction in the root-mean-square error depending on the cell line, a particularly exciting observation with regard to virtual screening problems where only a small subset of the predictions can be tested *in vitro*.

Extremely randomized tree

The extremely randomized tree (ERT) algorithm, another popular type of ensemble method, adds one further step of randomization than the random forest algorithm. The extreme randomness is introduced by the random subsampling of samples and features and random selection of features during branching. The randomness helps reduce variance to make the model more robust in dealing with noisy data, especially HTS data. Jeon's group presented a novel ERT-based drug combination discovery algorithm using various genomic information, drug targets and pharmacological information [35]. Genomic information included gene expression, mutation and CNV, and they selected only genes in cancer-related pathways to greatly reduce the feature dimension. Monotherapy and synthetic lethality information was also incorporated to improve prediction. They achieved a correlation coefficient of 0.738

Table 2. Complete list of published machine learning-based papers

Category	Algorithm	Cell line features	Drug features	Interaction	Synergy as feature	Dataset	Response	Cell line specific	Ref.
Traditional machine learning	Random forest	Drug-induced expression, Pathway information	Physicochemical properties, Drug-target information	Yes	No	DREAM	Classification	No	30
	Random forest	Gene expression, Mutation, CNV, Methylation, Gene–gene network, Monotherapy information	Drug-target information	Yes	No	DREAM	Regression classification	No	31
	XGBoost	Gene expression	Fingerprints, Binary toxicophore features	No	No	Merck & Co	Regression	No	32
	XGBoost	Gene expression, Mutation, CNV, Monotherapy information	Fingerprints, Drug-target information, Target protein domains, Targeted pathways	No	Yes	DREAM	Classification	No	33
	XGBoost		Fingerprints, MACCS keys, ISIDA/SIRMS fragments, Physicochemical properties	No	No	NCI-ALMANAC	Regression	Yes	34
	ERT	Gene expression, Mutation, CNV, Monotherapy information, Synthetic lethality	Drug-target information	No	No	Merck & Co	Regression Classification	No	35
	ERT		Drug-target information, Side effects, Drug structure, Gene ontology	No	No	NCI-ALMANAC	Classification	Yes	36
	Logistic regression	Gene expression, Gene essentiality, Pathway information	Drug-target information	No	No	NCI-ALMANAC	Classification	No	37
	Tensor factorization	Gene expression	Fingerprints	No	Yes	NCI-ALMANAC	Regression	No	38

Table 2. Continued

Category	Algorithm	Cell line features	Drug features	Interaction	Synergy as feature	Dataset	Response	Cell line specific	Ref.
Deep learning	Feed-forward neural network	Gene expression	Fingerprints Physicochemical properties Binary toxicophore features	No	No	Merck & Co	Regression Classification	No	39
	Feed-forward neural network	Gene expression MicroRNA expression Protein abundance	Fingerprints Molecular descriptors	No	No	NCI-ALMANAC	Regression	No	40
	Feed-forward neural network	Gene expression CNV Pathway information	Drug-target information	No	No	NCI-ALMANAC	Regression	No	41
	Tensor factorization Feed-forward neural network	Pathway information		No	Yes	Merck & Co	Classification	No	42
	Autoencoder	Gene expression Somatic mutation CNV	Fingerprints Physicochemical properties	No	No	Merck & Co	Regression	No	44
	End-to-end neural network	Gene expression Tissue/cancer type	Fingerprints SMILES representation Drug-target information	No	No	DrugComb	Regression Classification	No	45
	RBM	Gene expression Ontology fingerprints Pathway information	Drug-target information	Yes	No	DREAM	Classification	Yes	46
	GCN	Protein–protein interactions	Drug-target information	No	Yes	Merck & Co	Classification	Yes	48

There are two categories of methods—traditional machine learning and deep learning methods. The first three columns summarize algorithm, cell line and drug features. The 'Interaction' column indicates whether the method utilized the interaction between drug and cell line features as input to train the model. The 'Synergy as features' column indicates whether the model utilized known drug synergy information in the training set to construct drug synergy network features to be used for predicting unknown drug synergy. The 'Response' indicates the type of task each model chooses to solve. It can be a regression model, where the output target is a continuous synergy score, or a classification model, where the output target is a binarized label. The 'cell line specific' column indicates whether the model trained each cell line separately or as a whole. NCI, National Cancer Institute of the United States; Ref, study reference number.

between the predicted synergy scores and actual observations. Using the ERT model, they also generated synergistic rules from the frequent paths, and these rules can be further tested for biomarkers of drug combination therapies.

Gilvary's research team proposed a multi-task framework to study drug combination synergy in cells and in the clinic [36]. Assuming that each synergy metric may identify different types of synergistic combinations, they trained five models using different drug synergy measures (HSA, Bliss, Loewe, ZIP and ALMANAC score). Each model was a multi-task ERT framework in which each cancer cell line was considered as one task. No cell line features were required because the training process for each cell line was independent. The authors manually designed multiple pair-wise similarity features, including compound structure, side effects, drug targets, gene ontology and others, which were based on Pearson's correlation efficient and Jaccard and Dice similarity indices. They discovered that synergistic drug combinations exhibit lower distance between drug targets in the gene network compared to antagonistic drug pairs. The authors also found that each drug synergy metric identified unique synergistic drug pairs with distinct underlying joint mechanisms of action, which may shed light on the future design of a consistent drug synergy measure.

Logistic regression

Gene expression profiles, together with mutation and CNV, were always the first choice among cell line features for all the above-mentioned methods. However, gene essentiality, measured by clustered regularly interspaced short palindromic repeats or short-hairpin RNA, may also contribute to drug synergy prediction and merits investigation. Li *et al*. recently presented a study of how gene essentiality and pathway-level scores improve drug synergy prediction [37]. They employed logistic regression to test the statistical significance of each gene and pathway feature, comparing between gene expression and gene essentiality, and between targets and non-targets in their contribution to the prediction. They concluded that gene expression and essentiality exhibited different mechanisms and should be considered in combination rather than individually to improve the drug synergy prediction. Their team also observed that pathway features outperformed genes in the prediction, indicating that systems biology features can be used to boost prediction performance.

Tensor factorization

To capture the high-order interactions between drug combinations in different cell lines and at different doses, Julkenen's group proposed a matrix factorization method (comboFM) to model the multi-way interactions between two drugs, cell lines and dose–response matrices as fifth-order tensor data [38]. ComboFM can also integrate auxiliary drug and cell lines features, such as the molecular fingerprints of drugs and gene expression profiles of cancer cell lines, to aid prediction. Different from the above-mentioned machine learning models, comboFM first models drug combination effects at the level of the dose–response matrices and then quantifies the overall level of synergy of the combinations. Leveraging all the information contained in the dose–response matrices enables the learning of a more comprehensive view of the synergistic drug combination landscapes.

Deep learning methods

Feed-forward neural networks

The availability of large amounts of data and increasing computation power encourages and facilitates the utilization of deep learning methods to tackle the challenge of predicting anti-cancer drug synergy. DeepSynergy, the first deep learning-based model proposed to predict drug synergy, employed a three-layer feed-forward neural network architecture [39]. This model considered only gene expression as the cell line feature and three types of chemical descriptors (ECFP_6 fingerprints, physicochemical properties and binary toxicophore features) as drug features. It combined information about cell lines and drug combinations in its hidden layers to construct a combined hidden representation that eventually led to the accurate prediction of drug synergies. The far superior performance of DeepSynergy when compared with other traditional machine learning methods in a cross-validation setting indicates that deep learning is an ideal tool for solving the problem of drug synergy prediction.

Since the publication of DeepSynergy in 2018, several deep learning-based methods with similar architecture (feed-forward neural network) have been proposed. Feed forward is the most basic type of deep neural network, consisting of sequentially layered interconnected compute units (neurons). It can discover complex abstract structures in large datasets, employing a backpropagation algorithm to indicate how a machine should change the internal parameters it uses to compute the hidden representation in each layer from the representation in the previous layer [22]. Xia *et al*. presented a novel deep learning model to predict tumor cell line responses to drug combinations [40]. They considered gene expression, microRNA expression and protein abundance as cell line features and 30 categories of molecular descriptors and two types of fingerprints generated by Dragon software as drug features. Different from DeepSynergy, the researchers' model first applied feature-encoding sub-models to encode new vector representations for each feature type separately and then used the concatenation of these encoded features as input into a four-layer fully connected neural network model. Because the two drugs were symmetric, they shared the parameters of the encoding sub-model. The authors showed that the addition of the drug descriptors as features would improve the coefficient of determination, R^2, by 0.81, highlighting the informative value of drug features in the prediction of anti-cancer drug synergy.

Though the powerful representation capacity of deep neural networks supports their outstanding performance, the huge amounts of features and parameters in the existing architecture make it difficult to train and explain the model. To tackle this challenge, Zhang's research team introduced DeepSignalingSynergy, a novel simplified deep learning model to predict drug combination synergy in tumor cells [41]. Compared with existing methods that utilize a large number of chemical structure and genomics features in densely connected layers, construction of their model uses only a small set of informative cancer signaling pathways, thereby mimicking the integration of multi-omics data and drug features in a more biologically meaningful manner. In the first layer, they used gene expression, CNV and binary indicators of drug targets. In the second layer, each neuron corresponds with one signaling pathway, and the neurons in the first layer are connected to the neurons in the second layer only when the specific gene is included in the corresponding pathway. This results in a sparsely connected architecture, and the greatly reduced number of parameters needed for training leads to better generalization and interpretability. The authors also

used a layer-wise relevance propagation approach to investigate the importance of individual signaling pathways for prediction and showed that it was feasible to predict drug synergy based on a small set of informative signaling pathways.

Sun's group proposed a deep tensor factorization (DTF) algorithm for predicting anti-cancer drug synergy, which integrated a tensor factorization method and feed-forward deep neural network [42]. Unlike the previously mentioned machine learning methods, DTF used no cell line or drug features. Instead, DTF considered drug synergy data as multi-way data that can be best represented as a tensor. Tensor decomposition methods can be utilized to capture latent relationships between variables. DTF first used a constraint programming weighted optimization (CP-WOPT) algorithm to decompose tensors with missing entries and then used the results as input features to train a deep neural network model that predicts drug synergy. Performance was comparable between DTF and DeepSynergy, and DTF required far fewer training samples. However, this model's practical use is limited by its need for retraining by the addition of new entries to the tensor whenever a new drug or cell line needs to be considered.

Autoencoder

Besides basic feed-forward deep neural networks, autoencoder is another type of deep learning architecture that is used extensively to reduce the feature dimension by learning informative latent representations. Autoencoder is mainly designed to encode input features into a compressed and meaningful representation and then decode them back so that the reconstructed input is as similar as possible to the original [43]. Zhang's team proposed AuDNNsynergy, a novel deep learning model with autoencoders that integrates multi-omics and chemical structure data to predict drug synergy [44]. They separately trained three autoencoders for gene expression, mutation and CNV features using data of The Cancer Genome Atlas (TCGA) to transfer the knowledge embedded in the large-scale genomics data of the TCGA samples. The gene expression, mutation and CNV features of the cell line were then encoded using these trained autoencoders and were concatenated to be used as input for the prediction model. AuDNNsynergy used a feed-forward neural network as its prediction module and considered drug fingerprints and physicochemical properties as the input drug features as well as encoded multi-omics cell line features. The key contribution and innovation of this model is its combined use of a transfer learning approach with autoencoders to leverage the rich information contained in larger datasets.

Kim *et al.* also utilized transfer learning techniques to develop an autoencoder-based drug synergy prediction model to aid the study of data-poor tissues [45]. They aimed to utilize information from data-rich tissues that share some biological commonality in terms of gene expression with the understudied tissues and that might therefore respond similarly to drugs. The proposed architecture was an end-to-end multi-modal deep learning model consisting of feature-encoding (autoencoder) and prediction (feed-forward neural network) modules. In their feature-encoding module, cell line and drug encoders took raw features as input and output hidden representations. They considered gene expression and tissue/cancer type as cell line features and drug fingerprints, SMILES (simplified molecular input line entry system) representation, and drug-targeted genes as drug features. They employed a multi-layer feed-forward neural network in their prediction module that was similar to that of AuDNNsynergy. The model was pre-trained using data-rich tissues to enable

transfer of knowledge and then applied to predict drug synergy in understudied tissues. The main contribution of this study was to utilize transfer learning to investigate understudied but critical tissues for the highly accurate prediction of drug synergy.

Other architectures

Some other types of deep learning architecture have also been utilized to study anti-cancer drug synergy prediction. In order to integrate information from Ontology Fingerprints, gene expression and pathway information together, Chen *et al.* presented a stacked restricted Boltzmann machine (RBM) to predict the synergy of drugs [46]. The ability of the stacked RBM to combining unsupervised and supervised learning makes this deep belief networks favorable for the integration. The authors considered Ontology Fingerprints in this work which reflects literature information that could further improve drug synergy prediction. They created a RBM model for each cell line separately, and the input for each neuron of the RBM model is a quantified interaction of two drug-targeted genes.

In recent years, graph neural networks (GNNs) have become more and more popular for solving graph-related problems, and they have performed very well in many areas, including many applications in computational biology [47]. Because drug synergy prediction tasks can be modeled as link prediction problems in a drug–drug interaction network, Jiang *et al.* proposed a graph convolutional network (GCN) model to predict synergistic drug combinations in particular cancer cell lines [48]. For each cell line, they constructed a heterogeneous network consisting of three different types of subnetworks, a drug–drug synergy (DDS) network, DTI network and PPI network. The DDS network was based on the binarized synergy scores in the training set, and the researchers were trying to predict the missing edges of the DDS network that were contained in the test set. The GCN encoder was used to encode the heterogeneous network into a hidden embedding that was then decoded to infer potential synergistic drug pairs (edges in DDS network). This GCN-based model has been shown to achieve state-of-the-art performance and outperform DeepSynergy. However, the training of this model with very limited data regarding drug–protein interaction could have produced bias and thereby affected their prediction results.

Model comparisons

Prediction performances reported by each model in the original paper usually cannot be used to directly compare among models, as they typically utilized data from different databases or applied different filtering criteria. Therefore, in order to provide insightful information in terms of prediction capacity among reviewed models, we decided to conduct unbiased experiments using the same set of data for all models. We used NCI-ALMANAC dataset as the anti-cancer drug synergy data source [24]. NCI-ALMANAC contains data for 60 cell lines, and we only considered drugs with at least one target gene (68 drugs). In total, 130 182 samples were used for conducting comparisons. All NCI-60 cell line features (expression, mutation, CNV, etc.) were downloaded from CellMinerCDB [49]. Drug targets information were obtained from DrugBank [50], while drug molecular attributes were calculated using RDKit package in Python. We selected eight representative models to make comparisons, not including cell line-specific models or models based on tensor factorization. Hyperparameters used in these models were determined by 5-fold cross-validation. Since most of these models do not provide their source code, we implemented these models from scratch in

Table 3. Model prediction performance comparison results in terms of ROC-AUC and AUPR

Model	Algorithm	ROC-AUC	AUPR
32	XGBOOST	0.883 ± 0.005	0.357 ± 0.008
33	XGBOOST	0.891 ± 0.004	0.370 ± 0.014
35	ERT	0.849 ± 0.007	0.244 ± 0.013
37	Logistic Regression	0.846 ± 0.008	0.188 ± 0.007
39	Feed-forward neural network	0.913 ± 0.005	0.396 ± 0.015
40	**Feed-forward neural network**	$\mathbf{0.915 \pm 0.004}$	$\mathbf{0.429 \pm 0.019}$
44	Autoencoder	0.891 ± 0.005	0.356 ± 0.018
45	**End-to-end neural network**	$\mathbf{0.913 \pm 0.005}$	$\mathbf{0.423 \pm 0.021}$

'Model' column indicates the reference number of the model in the main text. The results in 'ROC-AUC' and 'AUPR' are in the form of mean $\pm$ SD from five repeated experiments. Highlighted rows are top-performing methods.

Python. Data were randomly split into 80% training set and 20% test set and were repeated five times to evaluate the model performance. All the simulations were carried out on a Linux system with Dual Intel Xeon 8268s CPU and Dual NVIDIA Volta V100 GPU. We used two evaluation metrics to compare the performance: area under the receiver operating characteristic curve (ROC-AUC) and area under the precision-recall curve (AUPR). It has been proved that AUPR is more informative than ROC-AUC when the labels are highly imbalance [51], which is our case. The detailed comparison results are shown in Table 3.

Deep learning methods outperform traditional machine learning methods

As we can see from the results in Table 3, among four traditional machine learning models and four deep learning models, almost all deep learning models outperform every traditional machine learning model in terms of both ROC-AUC and AUPR. Only XGBOOST algorithm can achieve comparable performance with deep learning models while ERT and logistic regression are a lot worse than deep learning models. The reason why deep learning methods outperform traditional machine learnings is that deep learning methods have greater ability to learn abstract representations from high-dimensional data, especially given large amounts of data. It can be imagined that with more and more data become available, the performance gap between deep learning methods and traditional machine learning methods will get wider.

How model processes features is more important than features themselves

Comparing between model [32] and model [33] where same model architectures are used while different input features are considered, we can notice little performance differences. However, using almost the same input features, the XGBOOST model outperforms the logistic regression model by a large margin in terms of AUPR, as logistic regression can only handle linear relationships between input features and labels while XGBOOST can deal with more complex non-linear relationships. This indicates the importance of the capacity of the model in processing input features when predicting anti-cancer drug synergy.

As for deep learning models, model [40] and model [45] achieve the best prediction performances. These two models both utilize feature encoding sub-modules to encode different types of input features separately (although they use different sets of input features), while DeepSynergy [39] first combines all types of input features and encodes them simultaneously. As for the autoencoder model [44], different from above three supervised models, it first utilizes unsupervised learning method (autoencoder) to obtain hidden representations and then uses the hidden representations as input features to a supervised neural network model. These four models differ greatly in the way the input features are processed, and we can conclude that defining sub-modules to process different types of features separately can help obtain the best performance when building deep learning models for anti-cancer drug synergy prediction. It is indicated that the way model processes features is more important than features themselves when it comes to the model prediction performance.

Discussion

Machine learning methods have been proven as powerful tools to tackle anti-cancer drug synergy prediction problems, especially as the availability of datasets has grown. They are able to derive predictive models without requiring strong assumptions about the mechanisms underlying the synergy. Given that simple linear models cannot accurately represent the complex nature of anti-cancer drug synergy, most published machine learning-based frameworks have employed ensemble or deep learning methods. XGBoost, ERT and random forest are popular traditional machine learning models, and feed-forward neural networks and autoencoder methods are favored deep learning models. These machine learning models have strongly outperformed such classic models as network pharmacology models with regard to prediction to enable the accurate prioritization of synergistic drug combination screenings.

Analysis of feature importance reveals potential underlying mechanisms

Machine learning methods can provide prediction results and elucidate feature contributions to facilitate interpretation. Though deep learning has been criticized for its black box nature, techniques such as the integrated gradients method have been proposed to explain the importance of input features [52]. Those genes or pathways assigned high importance values may indicate potential underlying mechanisms of drug synergy that can be further investigated. Comparison with cell line features has demonstrated the greater contribution of drug features to synergy prediction [32]. This may be attributable to the cell line features not being informative for drug synergy prediction or their not being used or encoded correctly utilizing current methods. Designing more effective and informative cell line features would be a key to further improve the performance of the prediction model.

Interactions between cell line and drug features

Most existing machine learning methods utilize cell line and drug features independently. However, all the original molecular cell line features, gene expression, mutation, CNV and methylation, are cell line specific and thus identical for samples from the same cell line but treated with different drug combinations, introducing redundant information. Apparently, different drugs play different roles in functional pathways, so the interaction between baseline molecular features and drug features can produce different post-treatment genomic information, and these

interaction features can provide much more meaningful information for prediction. Typically, it is difficult to obtain perturbation profiles for a large-scale pool of drug combinations. Therefore, these interaction features should be designed carefully, by either employing a network propagation algorithm to directly simulate post-treatment profiles based on a gene–gene network [31] or by calculating informative statistics based on both drug-target information and cell line-specific expression values [30, 46]. The fifth column of Table 2 summarizes whether the model utilized interaction features.

Curse of dimensionality

Existing machine learning models for predicting drug synergy use a large number of genomics and chemical features together, with input features numbering up to 10 000 or more. However, data sparsity and multicollinearity involving excessive numbers of features may greatly affect learning efficiency [53]. Besides, in a previously published review paper by DREAM Challenge [23], by comparing all submitted algorithms, they found that aggressive pre-filtering strategies have been successfully used to limit model complexity and improve model generalizability. Several approaches have been used to reduce the number of input dimensions, including filtering out genes with low variances [32] or considering only genes in cancer-related pathways [33, 35], differentially expressed genes [30] or active genes [46]. Another way to tackle this curse of dimensionality is to employ autoencoder models that can learn informative representations of lower dimension [44, 45], but the difficulty of interpreting the latent representations obtained by autoencoders does not favor this method for the assessment of biology-related tasks.

Future directions

Pioneering efforts have been made to develop machine learning models for the prediction of anti-cancer drug synergy, but some limitations remain to be addressed. The design of more system biology features that consider both drug target information and molecular profiles would improve prediction performance and biological interpretability. For example, instead of pathway-level analysis, simulation of the effects of drugs on cells at the sub-pathway level might provide more edifying and refined information to reflect the underlying signaling mechanisms. In addition, the design of frameworks that combine network pharmacology models, which explain mechanisms well, and machine learning models could aid prediction performance and interpretability. GNNs have the potential to integrate these two types of methodologies and merit further investigation, given its powerful ability to represent complex biological networks and its intrinsic nature of neural network models. Although all reviewed papers are proposed to study combinations of two drugs, it can be easy to extend existing machine learning methods to study combinations of multiple drugs. Taking DeepSynergy as an example, now the input features will be the concatenation of multiple drug features plus the cell line features instead of the concatenation of two drug features plus cell line features. However, in order for a specific multiple drug combinations synergy prediction model [54], special attention should be given to the model design such as solving the problem of symmetry and curse of dimensionality, which worth future investigations.

Another limitation of current models is the lack of ability to making concentration-dependent predictions. Most of the models use the average synergy scores under different concentrations as the target variable for the model training, losing concentration-specific information. However, it is important to know the combination efficacy specifically when each drug is administered at its clinically relevant concentration [55, 56]. There are two methods designed for predicting pre-clinical drug combination effects incorporating concentration information when developing models. IDACombo is an IDA-based method to predict the efficacy of drug combinations using monotherapy data based on the assumption that the expected effect of a combination of non-interacting drugs is simply the effect of the single most effective drug in the combination [57]. ComboFM is a novel machine learning framework for systematic modeling of drug-dose combination effects by factorizing higher-order tensors indexed by drugs, drug concentrations and cell lines [38]. More efforts should be made to study how to predict concentration-specific anti-cancer drug synergy.

Furthermore, the development of reliable machine learning models always suffers from insufficient training data. Lack of knowledge about drug combinations makes it difficult to gain new knowledge and make predictions. Using computational techniques to increase sample size is more efficient and less time-consuming and expensive than generating more data samples from experimental screening. One method to increase sample size and gain knowledge for drug synergy prediction would be to borrow information from the abundance of available drug sensitivity data. Multi-task or transfer learning can borrow information from the training signals of these data, which worth further investigation.

Key Points

- Greater availability of large-scale datasets and exponential growth in computing power have led to the increased popularity of machine learning methods.
- The choices of cancer cell line features and drug features are essential factors in developing machine learning models for anti-cancer drug synergy prediction.
- Machine learning methods are preferred in anti-cancer drug synergy prediction due to their hypothesis-free nature and outstanding predictive power.
- Ensemble methods and deep learning methods are the most popular machine learning algorithms in predicting anti-cancer drug synergy.

Availability of data and materials

The data and code used for comparison in this article are publicly available in Github, at https://github.com/kunjiefa n/anticancer-drug-synergy-prediction.

References

1. Jia J, Zhu F, Ma X, *et al*. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009;**8**:111–28.

2. O'Neil J, Benita Y, Feldman I, *et al*. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016;**15**:1155–62.

3. Eroglu Z, Ribas A. Combination therapy with BRAF and MEK inhibitors for melanoma: latest evidence and place in therapy. *Ther Adv Med Oncol* 2016;**8**:48–56.

4. Flaherty KT, Infante JR, Daud A, *et al*. Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N Engl J Med* 2012;**367**:1694–703.

5. Li X, Qin G, Yang Q, *et al*. Biomolecular network-based synergistic drug combination discovery. *Biomed Res Int* 2016;**2016**:1–11.

6. Ryall KA, Tan AC. Systems biology approaches for advancing the discovery of effective drug combinations Rajarshi Guha. *J Chem* 2015;**7**:1–15.

7. Bulusu KC, Guha R, Mason DJ, *et al*. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discov Today* 2016;**21**:225–38.

8. Tonekaboni SAM, Ghoraie LS, Manem VSK, *et al*. Predictive approaches for drug combination discovery in cancer. *Brief Bioinform* 2018;**19**:263–76.

9. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008;**4**:682–90.

10. Nelander S, Wang W, Nilsson B, *et al*. Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol* 2008;**4**:216.

11. Klinger B, Sieber A, Fritsche-Guenther R, *et al*. Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol Syst Biol* 2013;**9**:673.

12. Flobak Å, Baudot A, Remy E, *et al*. Discovery of drug synergies in gastric cancer cells predicted by logical modeling. *PLoS Comput Biol* 2015;**11**:1–20.

13. Yin N, Ma W, Pei J, *et al*. Synergistic and antagonistic drug combinations depend on network topology. *PLoS One* 2014;**9**:e93960.

14. Wu Z, Zhao XM, Chen L. A systems biology approach to identify effective cocktail drugs. *BMC Syst Biol* 2010;**4**:S7.

15. Xiong J, Liu J, Rayner S, *et al*. Pre-clinical drug prioritization via prognosis-guided genetic interaction networks. *PLoS One* 2010;**5**:e13937.

16. Cheng F, Kovács ÍA, Barabási AĹ. Network-based prediction of drug combinations. *Nat Commun* 2019;**10**:1197.

17. Cokol M, Chua HN, Tasan M, *et al*. Systematic exploration of synergistic drug pairs. *Mol Syst Biol* 2011;**7**:1–9.

18. Bishop CM. Pattern recognition and machine learning. 2006.

19. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. *Int J Remote Sens* 2007;**28**:823–70.

20. Amodei D, Ananthanarayanan S, Anubhai R, *et al*. Deep speech 2: end-to-end speech recognition in English and mandarin. *Int Conf Mach Learn* 2016;173–82.

21. Fan K, Guan Y, Zhang Y. Graph2GO: a multi-modal attributed network embedding method for inferring protein functions. *Gigascience* 2020;**9**:giaa081.

22. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–44.

23. Menden MP, Wang D, Mason MJ, *et al*. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun* 2019;**10**:1–17.

24. Holbeck SL, Camalier R, Crowell JA, *et al*. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res* 2017;**77**:3564–76.

25. Zagidullin B, Aldahdooh J, Zheng S, *et al*. DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019;**47**:W43–51.

26. Berenbaum MC. What is synergy? *Pharmacol Rev* 1989;**41**:93–141.

27. LOEWE S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 1953;**3**:285–90.

28. Yadav B, Wennerberg K, Aittokallio T, *et al*. Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Comput Struct Biotechnol J* 2015;**13**:504–13.

29. Liu H, Zhang W, Zou B, *et al*. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020;**48**:D871–81.

30. Li X, Xu Y, Cui H, *et al*. Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif Intell Med* 2017;**83**:35–43.

31. Li H, Li T, Quang D, *et al*. Network propagation predicts drug synergy in cancers. *Cancer Res* 2018;**78**:5446–57.

32. Janizek JD, Celik S, Lee S-I. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. *bioRxiv* 2018;331769.

33. Celebi R, Bear Don't Walk O, Movva R, *et al*. In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Sci Rep* 2019;**9**:1–10.

34. Sidorov P, Naulaerts S, Ariey-Bonnet J, *et al*. Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front Chem* 2019;**7**:1–13.

35. Jeon M, Kim S, Park S, *et al*. In silico drug combination discovery for personalized cancer therapy. *BMC Syst Biol* 2018;**12**:16.

36. Gilvary CM, Dry JR, Elemento O. Multi-task learning predicts drug combination synergy in cells and in the clinic. *bioRxiv* 2019;576017.

37. Li J, Huo Y, Wu X, *et al*. Essentiality and transcriptome-enriched pathway scores predict drug-combination synergy. *Biology (Basel)* 2020;**9**:1–18.

38. Julkunen H, Cichonska A, Gautam P, *et al*. Leveraging multiway interactions for systematic prediction of pre-clinical drug combination effects. *Nat Commun* 2020;**11**:6136.

39. Preuer K, Lewis RPI, Hochreiter S, *et al*. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46.

40. Xia F, Shukla M, Brettin T, *et al*. Predicting tumor cell line response to drug pairs with deep learning. *BMC Bioinformatics* 2018;**19**:486.

41. Zhang H, Feng J, Zeng A, *et al*. Predicting tumor cell response to synergistic drug combinations using a novel simplified deep learning model. *AMIA Annu Symp Proc* 2020;**2020**:1364–72.

42. Sun Z, Huang S, Jiang P, *et al*. DTF: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics* 2020;**36**:4483–89.

43. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;**61**:85–117.

44. Zhang T, Zhang L, Payne P.R., *et al*. Synergistic drug combination prediction by integrating multiomics data in deep

learning models. *Translational Bioinformatics for Therapeutic Development* 2021; Humana, New York, NY, 223–38.

45. Kim Y, Zheng S, Tang J, *et al*. Anti-cancer drug synergy prediction in understudied tissues using transfer learning. *Journal of the American Medical Informatics Association* 2021;**28**:42–51.

46. Chen G, Tsoi A, Xu H, *et al*. Predict effective drug combination by deep belief network and ontology fingerprints. *J Biomed Inform* 2018;**85**:149–54.

47. Nelson W, Zitnik M, Wang B, *et al*. To embed or not: network embedding as a paradigm in computational biology. *Front Genet* 2019;**10**:1–11.

48. Jiang P, Huang S, Fu Z, *et al*. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Comput Struct Biotechnol J* 2020;**18**:427–38.

49. Luna A, Elloumi F, Varma S, *et al*. CellMiner cross-database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res* 2021;**49**:D1083–93.

50. Wishart DS, Feunang YD, Guo AC, *et al*. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.

51. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*, 2006. pp. 233–40.

52. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. *Int Conf Mach Learn* 2017; 3319–28.

53. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. 2001;**1**.

54. Ianevski A, Giri AK, Aittokallio T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res* 2020;**48**:W488–93.

55. Zhou Y, Wang F, Tang J, *et al*. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Heal* 2020;**2**: e667–76.

56. Zhou Y, Hou Y, Shen J, *et al*. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* 2020;**6**:1–18.

57. Ling A, Huang RS. Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action. *Nat Commun* 2020;**11**: 5848.