COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Review

Machine learning in the prediction of cancer therapy

Raihan Rafique [a,1], S.M. Riazul Islam [b,1], Julhash U. Kazi [c,d,*]

[a] *Ideflod AB, Lund, Sweden*
[b] *Department of Computer Science and Engineering, Sejong University, Seoul, South Korea*
[c] *Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University, Lund, Sweden*
[d] *Lund Stem Cell Center, Department of Laboratory Medicine, Lund University, Lund, Sweden*

A R T I C L E   I N F O

A B S T R A C T

Resistance to therapy remains a major cause of cancer treatment failures, resulting in many cancer-related deaths. Resistance can occur at any time during the treatment, even at the beginning. The current treatment plan is dependent mainly on cancer subtypes and the presence of genetic mutations. Evidently, the presence of a genetic mutation does not always predict the therapeutic response and can vary for different cancer subtypes. Therefore, there is an unmet need for predictive models to match a cancer patient with a specific drug or drug combination. Recent advancements in predictive models using artificial intelligence have shown great promise in preclinical settings. However, despite massive improvements in computational power, building clinically useable models remains challenging due to a lack of clinically meaningful pharmacogenomic data. In this review, we provide an overview of recent advancements in therapeutic response prediction using machine learning, which is the most widely used branch of artificial intelligence. We describe the basics of machine learning algorithms, illustrate their use, and highlight the current challenges in therapy response prediction for clinical practice.

Contents

* Corresponding author at: Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University, Medicon village Building 404:C3, Scheelevägen 8, 22363 Lund, Sweden.
  E-mail address: kazi.uddin@med.lu.se (J.U. Kazi).
[1] Equal contributions.

1. Introduction

Adaptive resistance mechanisms are highly dependent on cancer subtypes and applied treatments. Therefore, the resistance mechanism needs to be defined for each cancer subtype and individual treatment plan. Currently, hardly any tools exist to determine from the beginning whether a patient will respond to a specific therapy or display resistance. Thus, there is an unmet need to develop tools to identify drug responses in individual patients for precision medicine. Recent technological advances have initiated a new era of precision medicine through data-driven assessment of diseases by combining machine learning (ML) and biomedical science. The use of artificial intelligence such as ML helps to extract meaningful conclusions by exploiting big data, thereby improving treatment outcomes. ML is widely used in cancer research and is becoming increasingly popular for cancer detection and treatment. The main goal of precision medicine is to provide therapies that not only increase the survival chances of patients but also improve their quality of life by reducing unwanted side effects. This can be achieved by matching patients with appropriate therapies or therapeutic combinations.

Some of the early studies on ML and its applications in human cancer research have been discussed elsewhere [1]. Several recent overviews in this emerging field have provided valuable insights into the relevant computational challenges and advancements [2–8]. These overviews illustrated the importance of the field and supported the notion that ML is a highly promising approach to personalized therapy for cancer treatment. In a recent review, a broad perspective was provided on how ML tools can be incorporated into clinical practice with a focus on biomarker development [9]. Another review identified several challenges in omics data analysis and data integration to obtain robust results in big-data-assisted precision medicine [10]. Several other reviews dealt primarily with the computational methods and software that are required to advance data-driven precision oncology [11–13]. Also, whereas Grothen et. al. discussed artificial intelligence-based investigations into cancer subtypes and disease prognosis from a system biology perspective [14], Biswas et. al. reviewed artificial intelligence applications for pharmacy informatics in a surveillance and epidemiological context [15]. Another study systematically explained how deep learning (DL), a subset of ML, has emerged as a promising technique, highlighting various genomics and pharmacogenomics data resources [16]. However, the aforementioned studies did not focus strictly on drug response prediction from clinical perspectives. In recent years, several surveys and review articles have presented the potential and challenges of ML adoption in clinical practice and drug response prediction in cancer treatment [17–23]. Nonetheless, the area of applications of ML in cancer treatment is so diverse that various issues still need to be analyzed from a holistic perspective. In this review, we provide a comprehensive overview of the ML solutions for drug response prediction relating to the relevant clinical practices. In addition to discussing the basics of therapy response prediction and related ML principles, we systematically present the ML and DL approaches that are promising for monotherapy and combination therapy in cancer treatment, a focus that makes our article different from existing surveys and reviews.

2. Basics of therapy response prediction

Predictive model development involves several steps that combine biological data and ML algorithms. A brief workflow has been depicted in Fig. 1.

2.1. Pharmacogenomic data resources

High-quality biological data are a prerequisite for a good model. Large-scale cell line data are publicly available from different platforms and include genomic, transcriptomic, and drug response data. Pharmacogenomic data for cell lines are available mainly from the Cancer Cell Line Encyclopedia (CCLE) [24,25], NCI-60 [26], the Genomics of Drug Sensitivity in Cancer (GDSC) [27,28], gCSI [29], and the Cancer Therapeutics Response Portal (CTRP) [30,31]. PharmacoDB [32] and CellMinerCDB [33,34] provide access to the curated data from different studies. These datasets offer baseline genomic and transcriptomic data for cell lines covering a wide range of cancers. DrugComb [35] and DrugCombDB [36] offer manually curated drug combination data from different
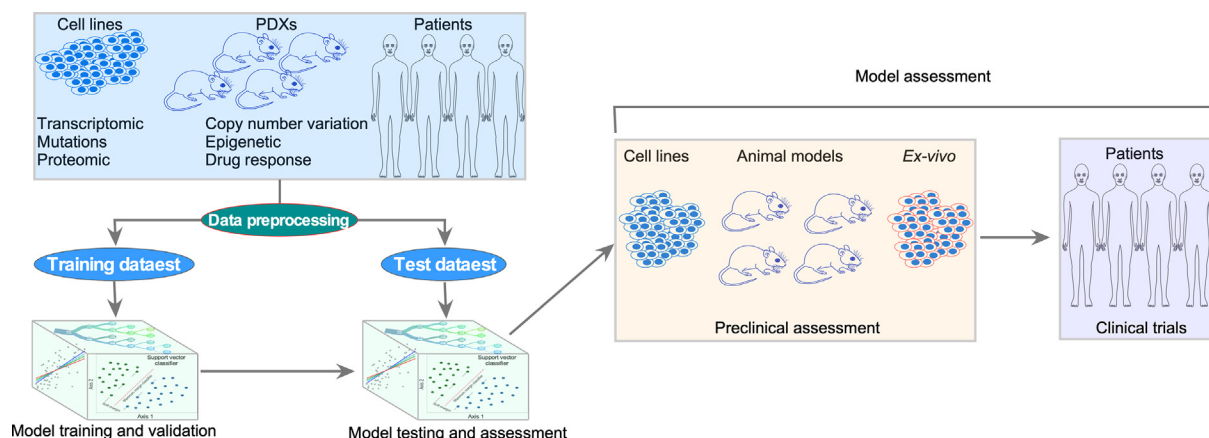
Fig. 1. Workflow for ML prediction model development. Pharmacogenomic data from cell lines, patient-derived xenografts (PDXs), and patient materials are ideal for ML model development. Data from different sources are preprocessed and then divided into training (including cross-validation) and test groups. The training dataset is used to build and validate the prediction model, while the test dataset is used for testing the model's accuracy and precision. To develop a prediction model for clinical use, vigorous preclinical assessment is required that can be performed using cell lines, PDXs, and patient materials that have not been used for model development. Additionally, the efficacy of predicted drugs must be tested for disease-specific preclinical models. Finally, both the model and predicted drug will undergo a clinical trial.

studies. Besides these pharmacogenomic data for cell lines, which have been widely used to develop ML models, several initiatives have recently been undertaken to generate pharmacogenomic data from patient-derived xenografts (PDXs). Compared with cell lines, PDXs are superior in predicting clinical activities. PDX finder [37], PRoXE [38], PDMR [39], and EorOPDXs [40] provide comprehensive data for PDXs. Several other studies also provide high-quality transcriptomic and pharmacogenomic data that are useful for model development or testing when combined with other datasets [41–45].

2.2. Data preprocessing

Data preprocessing is an important step in the ML approach. Large-scale data preprocessing includes data selection, noise filtering, imputation of missing values, feature selection, and normalization.

Data selection – Data selection remains the most challenging aspect due to the possible inconsistencies between different datasets [46]. Studies comparing the largest public collections of pharmacological and genomic data for cell lines suggest that each dataset separately exhibits reasonable predictive power but that combining datasets can further increase the classification accuracy [29,47].

Feature selection – Large-scale cell line datasets comprise transcriptomic, mutational, copy number variation (CNV), methylation, and proteomic data. Although genetic features such as mutations, CNV, and promotor methylation have been shown to provide important therapeutic insights, these features seem to be limited to individual tumors [27]. Therefore, it has been suggested that transcriptomic features alone hold the most predictive power and that the addition of genetic features marginally improves performance of an ML model [48–50]. The feature-to-sample ratio plays an important role in controlling the variances, with a smaller ratio providing better prediction [51]. However, maintaining a proper feature-to-sample ratio is challenging for pharmacogenomic data. For example, transcriptomic data can have more than 15,000 features, while the number of samples in any pharmacogenomic study remains between 100 and 1000. Systematically reducing the number of features (also known as dimensionality reduction) by incorporating meaningful descriptions improves prediction accuracy by reducing overfitting [52,53]. Several techniques can be used for feature selection, including minimum

redundancy maximum relevance (mRMR), high-correlation filters, principal component analysis, and backward feature elimination [54–62].

Data normalization – Because the range of values of raw data varies widely, a normalization technique (also known as feature scaling) is applied to change the values of numeric columns in the dataset to obtain a common scale, so that the associated objective functions work properly. Different ways exist to perform feature scaling, including min–max normalization, rank-invariant set normalization, data standardization, cross-correlation, and scaling to unit length [63].

3. ML algorithms for drug response prediction

ML algorithms can be grouped into four major classes: supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning [64,65]. Supervised learning algorithms use a training dataset with known outcomes to build a hypothetical function with decision variables that can later be used to predict unknown samples (Fig. 2). On the other hand, unsupervised learning algorithms use unlabeled data to find hidden structures or patterns; these algorithms are widely used in biological research for clustering and pattern detection. Semi-supervised learning algorithms are self-learning and can develop a prediction model from partially labeled data [66]. A reinforcement learning algorithm employs a sequential decision problem in which the algorithm solves a problem and learns from the solution [65]. In this case, the algorithm discovers which actions result in the best output on a trial-and-error basis. Perhaps supervised learning algorithms are generally used for building classification models, and these algorithms have also been widely tested for predicting treatment outcomes. Therefore, in this review, we will focus mainly on supervised learning algorithms.

3.1. Linear regression

Linear regression algorithms are simple and constitute the most popular ML algorithms, with a wide range of applications. The standard algorithm, least squares regression, uses the sum of squared residuals as the cost function to be minimized. Least squares regression works with a simple dataset; however, with increasing complexity, the algorithm shows overfitting (low bias but large variance). To resolve this problem, several algorithms,
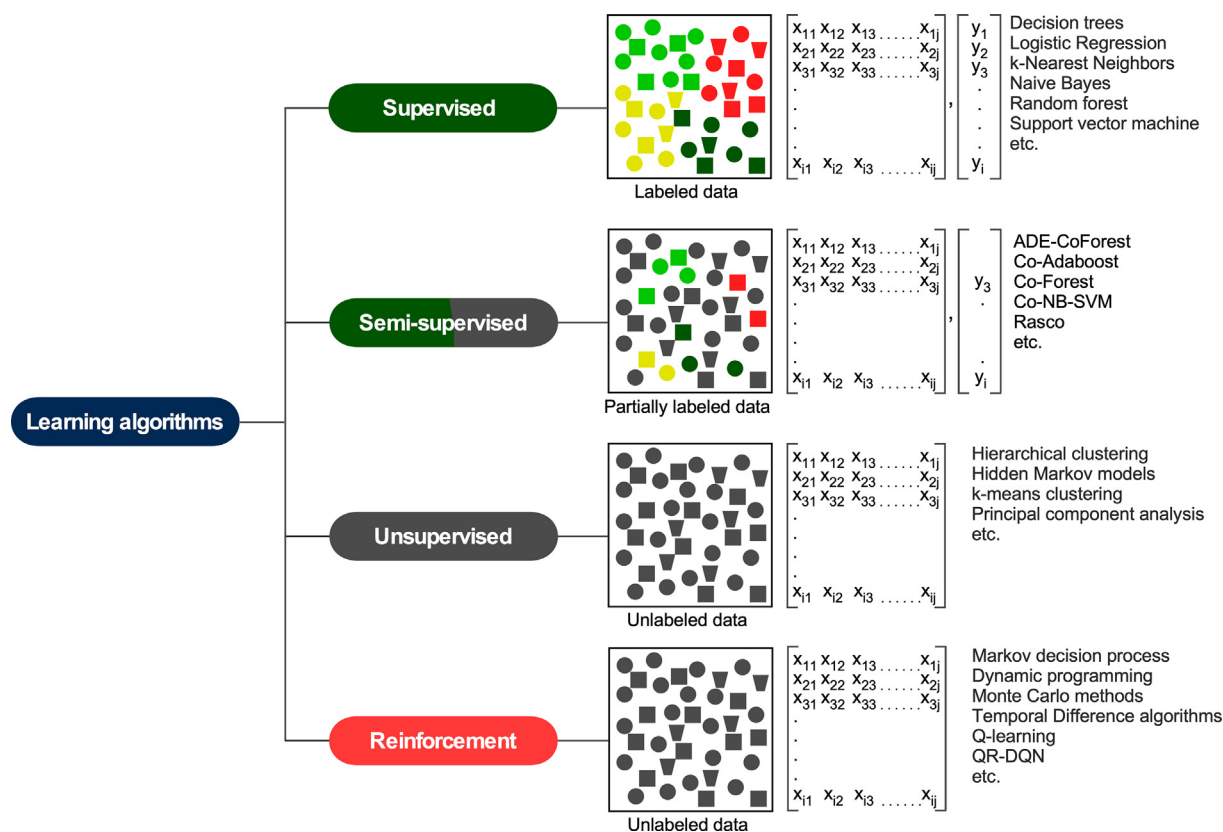
Fig. 2. Schematic representation of different ML algorithms. In a supervised learning model, all data have a known label, while the semi-supervised model can handle partially labeled data. Both unsupervised and reinforcement learning algorithms can handle unlabeled data.

such as the ridge model, lasso model, and elastic net, have been proposed. The cost functions in these models have been modified to increase the bias and reduce the variance. In a ridge model, a so-called L2 regularization, which is the squared value of the slope multiplied by λ, has been added to the least squares cost function. The least absolute shrinkage and selection operator (lasso) regularization (known as L1 regularization) is similar to the ridge regularization, but in this case, the added value is the absolute value of the slope multiplied by λ. The elastic net algorithm adds contributions from both L1 and L2 regularization; the cost function = min (sum of the squared residuals + λ * squared value of slope + λ * absolute value of slope). The λ parameter is a positive number that represents regularization strength. A larger λ value specifies stronger regularization, while a near-zero value removes the regularization so that all three algorithms become similar to the least squares model (Fig. 3). By changing the value of λ, it is possible to select meaningful features. Therefore, these methods can be applied to feature selection as well as to classification and regression problems [24,28].

3.2. Nonlinear regression

Among the various supervised learning algorithms, the decision tree is a relatively popular predictive modeling algorithm used to classify simple data. A decision tree takes data in the root node and, according to a test rule (representing the branch), keeps growing until it reaches a decision (representing a leaf node). The internal nodes represent different attributes (features) [67]. Each internal node breaks the data into a small subset until it meets a particular condition. It is a white-box-type algorithm, as each step can be understood, interpreted, and visualized. Although the decision tree is useful for simple classification, with a larger dataset

that has many features, it displays poor prediction powers due to overfitting. To resolve this problem, several advanced decision-tree-based models have been developed. The random forest algorithm randomly splits (bootstrapping) training data into several subsets (bagging) and uses each subset to build decision trees (Fig. 4). The use of multiple random decision trees for prediction increases the prediction accuracy [68]. Apart from the parallel use of random multiple decision trees, boosting algorithms, such as adaptive boosting (AdaBoost) and gradient boosting, use decision trees sequentially [69,70]. AdaBoost usually uses one-node decision trees (decision stump), while gradient boosting uses decision trees of between 8 and 32 terminal nodes. Both adaptive and gradient boosting algorithms display better prediction performance than single decision trees. Furthermore, a more regularized gradient boosting algorithm, extreme gradient boosting (XGBoost), outperforms the former gradient boosting algorithms [71].

3.3. Kernel functions

Kernel functions are widely used to transform data to a higher-dimensional similarity space. Kernel functions can be linear, non-linear, sigmoid, radial, polynomial, etc. Support vector machines (SVMs) are among the most popular kernel-based algorithms that can be used not only for supervised classification and regression problems but also for unsupervised learning. In a two-dimensional space, a linear SVM classifier is defined by a straight line as a decision boundary (maximum margin classifier) with a soft margin (Fig. 5A). In this case, the soft margins are also straight lines that represent the minimal distance of any training point to the decision boundary [72]. With simple one-dimensional data, the decision boundary can be a point (Fig. 5B); however, for com-
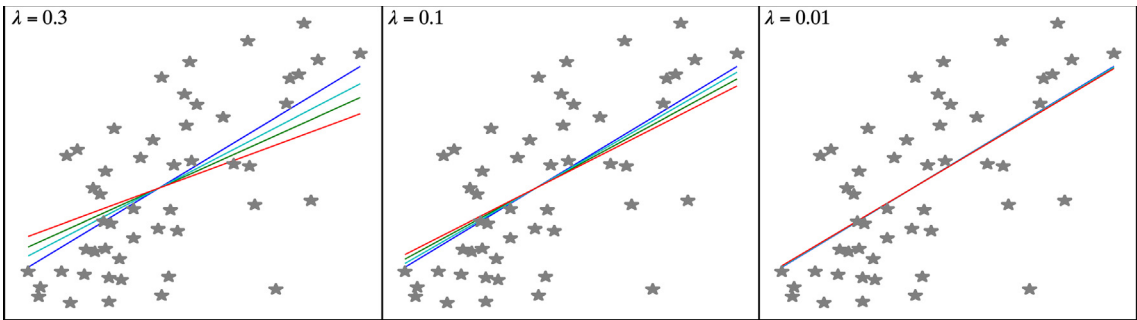
Fig. 3. A comparison of different linear regression algorithms. The sklearn.linear_model from SciKit learn was used to generate example plots using a diabetes dataset provided in SciKit learn. Plots show that by changing the λ value, regression can be regulated such that with a small λ value, all linear regression algorithms provide similar regression. Color code: linear regression – blue, ridge regression – green, lasso – cyan, and elastic net – red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
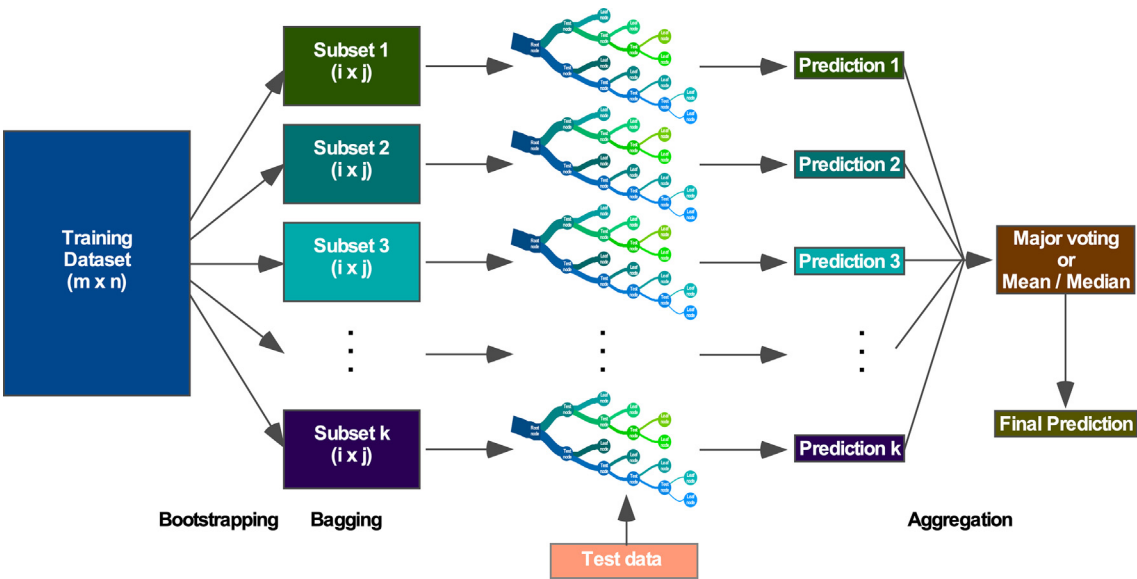
Fig. 4. Schematic representation of random forest algorithm. The three major steps in the random forest algorithm are bootstrapping, bagging, and aggregation. During bootstrapping, the training dataset is resampled into several small datasets, which are then bagged for the decision tree. The size of the bagged dataset remains the same but bootstrapped decision trees are different from each other. All decision trees make predictions on test data, and in the aggregation step, all predictions are combined for the final prediction. For a classification problem, the final prediction is made by major voting, but for a regression problem, the final prediction uses the mean or median value.
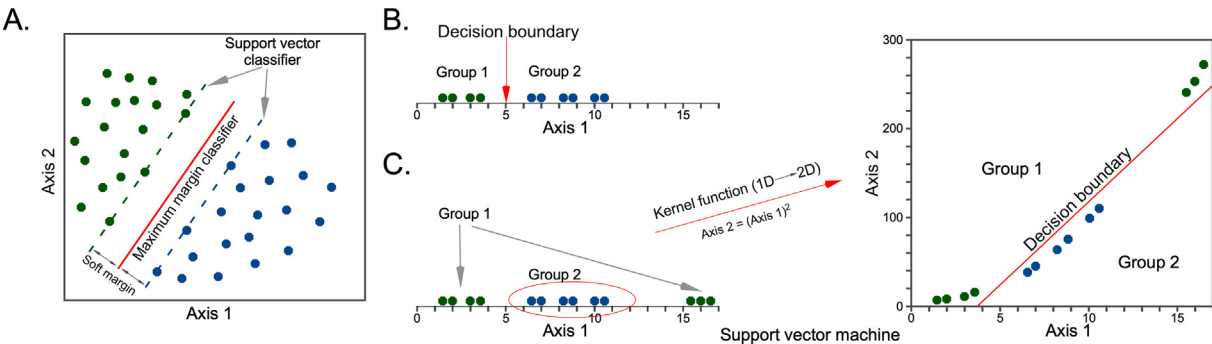
Fig. 5. Support vector machine. (A) In a two-dimensional SVM classification system, the maximum margin classifier is a straight line (red line). Support vectors are the nearest data points from the maximum margin classifier. The distance between support vectors and the maximum margin classifier is denoted as the soft margin. (B) In a two-group, one-dimensional data space, the decision boundary is a point, as shown by the red line. (C) In a two-group one-dimensional data space where the decision boundary cannot be drawn by a point, data are transformed by a kernel function to increase the dimension. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plex problems, the data may need to be transformed to a higher dimension to draw a decision boundary (Fig. 5C).

3.4. Deep learning

DL methods are a type of ML method that can automatically discover appropriate representations for regression or classification problems upon being fed with suitable data. The model can learn complex functions and amplify important aspects to suppress irrelevant variations. During training, the algorithm takes the raw input and processes it through hidden layers using nonlinear activation functions. The algorithm tries to minimize certain cost functions by defining values for the weights and biases (Fig. 6A). Usually, gradient descent is used to find the minima. Gradients for all modules can be determined by using the chain rule for derivatives, a procedure that is known as backpropagation (starting from the output and moving toward the input) [73]. DL algorithms have been successfully employed in various domains, including image classification, because of the availability of more data than features. The development of DL models using genomic or transcriptomic data is challenging due to the limited number of samples and the presence of many features. The selection of appropriate features can reduce the feature-to-sample ratio and, thereby, prevent overfitting. Furthermore, the addition of random dropout layers can help the model learn important features and reduce overfitting (Fig. 6B).

Convolutional neural networks (CNNs) are useful for feature learning (Fig. 6C). During the convolution and pooling steps, the algorithm of a CNN learns important features [73]. CNNs are widely used for structured data, such as images; however, if the data are stored in other types of architectures, such as graphs (an example includes small-molecule drugs with multiple atoms and chemical bonds), conventional CNNs cannot be used. In this case, a different type of convolutional neural network, referred to as the graph convolutional networks (GCNs), could be applied to the graph data [74]. GCNs have especially been used to extract atomic features from drug structure (graph) data [75].

4. Monotherapy response prediction

Currently, only a few drug response prediction tools are available for clinical use. In fact, a couple of linear regression prediction models are currently being used for certain types of cancers. A supervised classification model using a 70-gene signature was developed in 2002 to predict chemotherapy responses in breast cancer [76]. The method was patented as MammaPrint and is currently used in the clinic for patients with early-stage breast cancer. Later, a similar method was developed in which a linear regression model based on the scores of a 21-gene signature (Oncotype DX) was used to predict the chemotherapy responses in early-stage, estrogen-receptor-positive, HER2-negative invasive breast cancer [77]. Furthermore, a 50-gene signature was employed in multivariate supervised learning (PAM50 or Prosigna, a breast cancer prognostic gene signature assay) to predict treatment responses in breast cancer [78]. Aside from these simple, cancer-subtype-specific prediction models that are currently available in the clinic, most other studies regarding monotherapy predictions are still in the preclinical phase. Fig. 7 shows an overview of the methods that have been used to develop monotherapy prediction models in the past decade (a brief overview is included in Table 1).
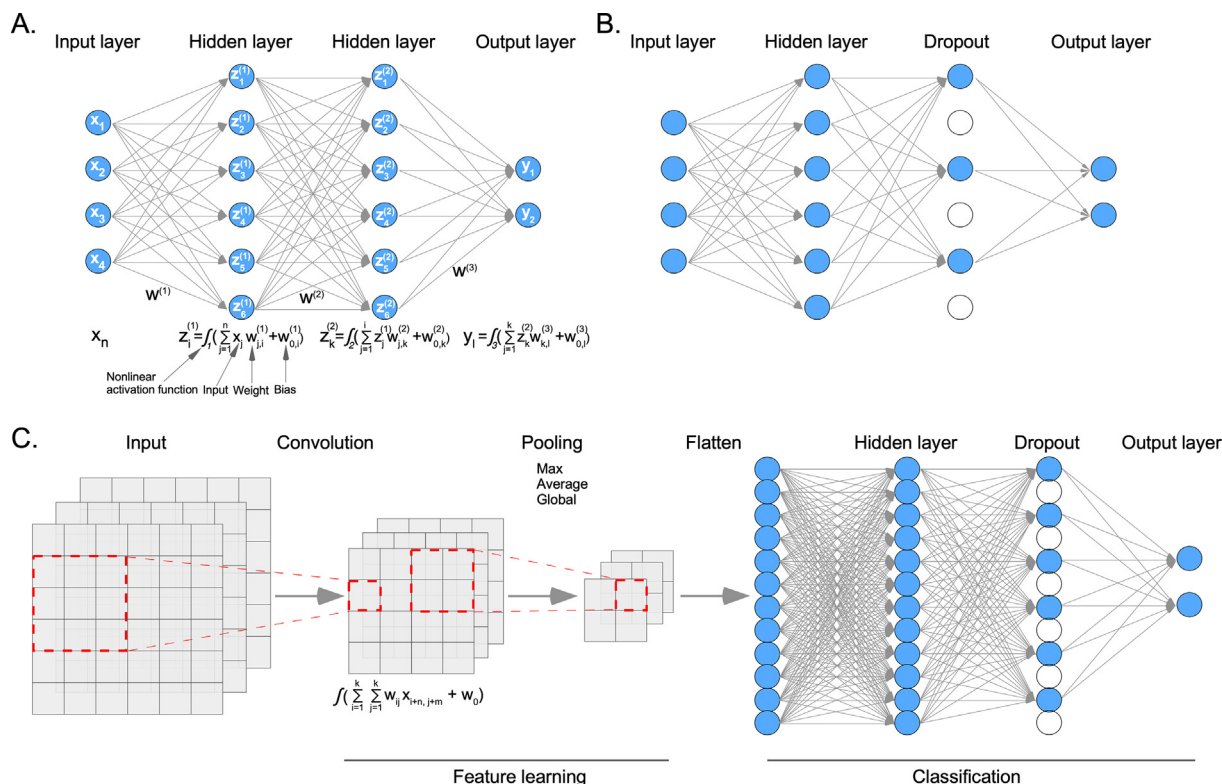
Fig. 6. Deep learning (DL). (A) In a deep neural network (DNN) model, each node of the input data layer is fully connected to the hidden layer nodes. The first hidden layer takes input data, multiplies it by weight, and adds a bias before applying a nonlinear activation function. The second hidden layer takes the first hidden layer as input and so on until it reaches the output layer. (B) In a dropout layer, some nodes are randomly removed. (C) During the convolution, the dimension of input data is reduced using a certain kernel size (in this example, 3x3) and the activation function. Then, features are pulled for further reduction. Finally, pulled features are flattened and applied to a DNN.
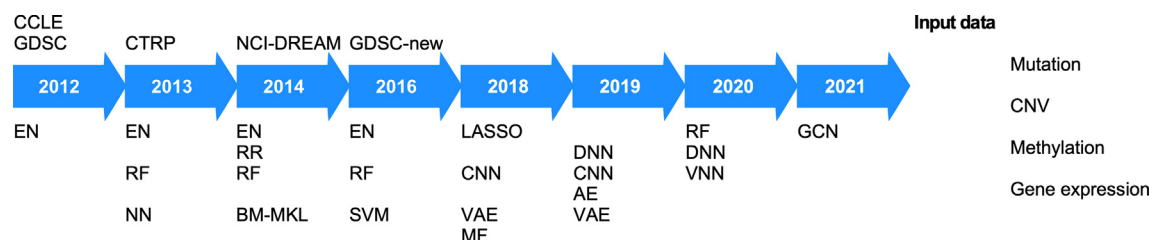
Fig. 7. ML algorithms used in the last decade to build monotherapy response prediction. Earlier prediction models were likely developed mainly using classical ML algorithms. Later, the DL algorithms were used mostly to develop the models. The majority of the studies used multi-omics data (mutation, CNV, methylation, and gene expression) collected from large screening studies such as CCLE, GDSC, CTRP, etc. EN – elastic net, RF – random forest, NN – neural network, RR – ridge regression, BM-MKL – Bayesian multitask multi-kernel learning, SVM – support vector machine, LASSO - least absolute shrinkage and selection operator, CNN – convolutional neural network, DNN – deep neural network, AE – autoencoder, VAE – variational autoencoder, MF – matrix factorization, VNN – visual neural network, GCN – graph convolutional network.

Table 1
Studies predicting monotherapy responses.

| Year	Data	Features	Algorithm	Ref.
2012	GDSC	Mutation, CNV, gene expression	Elastic net	[28]
	CCLE	Mutation, CNV, gene expression	Elastic net	[24]
2013	CCLE, GDSC	Gene expression (1000 selected genes)	Elastic net and other	[54]
	CTRP	Mutation, CNV	Elastic net	[30]
	GDSC	Selected genomic features	Neural networks and random forests	[80]
2014	GDSC, clinical data	Gene expression	Ridge regression	[79]
	CCLE, GDSC	Mutation, CNV, gene expression	Elastic net and ridge regression	[50]
	GDSC, CCLE, NCI	Gene expression (1000 selected genes)	Random forest	[55]
	NCI-DREAM	Mutation, CNV, gene expression, proteomic	BM-MKL	[49]
2015	GDSC, CCLE	Gene expression	Cell line-drug network model	[82]
2016	NCI	Mutation, CNV, gene expression, RPLA, miRNA	Random forest and support vector machine	[81]
	GDSC 2	Mutation, CNV, gene expression, methylation	Elastic net and random forest	[27]
	LINCS	Gene expression	DNN	[88]
2018	AML patient and cell line data	Gene expression	VAE + LASSO (DeepProfile)	[99]
	GDSC	Genomic fingerprints	CNN	[91]
	AML patient and cell line data	Gene expression, mutation, CNV, methylation	Network-based gene-drug associations	[87]
	PharmacoDB, CMap	Gene expression	VAE (Dr.VAE)	[59]
	CCLE, GDSC	Gene expression	Recommender systems	[94]
2019	GDSC	Gene expression	DNN	[90]
	TCGA, CCLE	Mutation, gene expression	VAE, DL (DeepDR)	[60]
	GDSC	Mutations and CNV	CNN ((tCNNS)	[105]
	GDSC	Mutation, CNV, gene expression.	DL (MOLI)	[92]
	GDSC, CCLE	Gene expression	Autoencoder (DeepDSC)	[61]
2020	PDXGEM	Gene expression	Random forest	[106]
	GDSC, KEGG, STITCH	Gene expression, pathway	DL	[89]
	GDSC, CCLE, CTRP	Gene expression, mutation, CNV, methylation	VNN	[62]
	van de Wetering et al. [108], Lee et al. [109]	Gene expression, pathway	Ridge regression	[107]
2021	GDSC	Mutations and CNV	Graph convolutional network	[104]

4.1. Classical ML models in monotherapy prediction

Sparse linear regression models have been used to predict drug sensitivity in initial large-scale pharmacogenomic studies with cell lines from various cancers [24,28,30]. These studies combined genomic features with transcriptomic features from cell lines and correlated them with corresponding drug sensitivity scores. The ridge regression and elastic net algorithms were predominantly employed for predictions [24,28,30,50,79]. However, due to the linear nature of the algorithms and the use of many features, these models could easily become overfitted.

As discussed above, the performance of prediction algorithms is largely influenced by biological feature selection [54,55,80,81]. Prediction performance can further be improved by incorporating information on the similarity between cell lines and drugs [82]. Cell lines with a similar gene expression profile show similar responses to a specific drug, while drugs with a similar chemical structure display similar inhibitory effects toward different cell lines. Therefore, a dual-layer network model that also considers similarity information outperforms linear models [82]. Likewise, a method based on a heterogeneous network in which the relationships among drugs, drug targets, and cell lines were explicitly

incorporated was shown to better capture the relationship between cell lines and drugs [83]. Collectively, a predictive model with selected features performs better, and the addition of network features improves the prediction accuracy.

The community-based NCI-DREAM study used a limited number of samples with a large number of genomic, transcriptomic, and proteomic features [49]. The NCI-DREAM initiative developed 44 different drug sensitivity prediction models, with the Bayesian multitask multikernel learning (BM-MKL) models performing relatively better than other models. BM-MKL includes Bayesian inference, multitask learning, multiview learning (multiple data view), and kernelized regression [49,84,85]. The standard model, kernelized regression, is a nonlinear classification algorithm similar to SVMs. Unlike the elastic net, kernelized regression captures the nonlinear relationship between drug sensitivity and genomic or transcriptomic features but simplifies the process by using a single component for the predictions.

Besides using genomic or transcriptomic features to predict drug sensitivity, the chemical and structural properties (also known as descriptors) of drugs have been incorporated into the learning algorithms. Combining drug descriptors with genomic or transcriptomic data allows for the simultaneous prediction of

multiple drug responses from a single model, although it is a challenging task due to the further increase in the total number of features [86]. Likewise, in a study with multicancer and multidrug associations, a disease-specific multi-omics approach to predicting gene-drug association was adopted in which each gene was checked for a pathway association [87]. The method is useful for identifying critical regulatory genes that can be targeted by a drug.

4.2. Deep neural networks in monotherapy prediction

Although DL has long been widely used in several areas of medical science and drug discovery platforms, it has recently been applied to drug response prediction as well. Initially, feedforward deep neural networks (DNNs) were applied to develop models using selected genomic features [80] or transcriptomic data [88]. Later studies incorporated selected gene expression features with pathway information to build DNN models [89,90]. In any case, all these DNN models have been shown to outperform classical ML models.

A CNN was used in the Cancer Drug Response Profile scan (CDRscan) study, in which convolutions were applied separately to genomic fingerprints of cell lines and molecular fingerprints of drugs [91]. After convolution, those two sets of features were merged and used with the drug response data to develop a DNN model. Because a CNN learns important features during training [73], the CDRscan method displays considerably higher robustness and generalizability. A similar model (MOLI) was developed using somatic mutations, CNVs, and gene expression data from GDSC [92]; the model was later validated with PDXs and patient samples.

4.3. Matrix factorization and factorization machines in monotherapy prediction

Matrix factorization (MF) is a supervised learning method that has been widely used in popular e-commerce ML recommender systems [93]. MF takes high-dimensional data, with missing information, as input and decomposes it into lower-dimensional matrices with the same numbers of latent factors (Fig. 8A). The learning algorithms in recommender systems are not general and must be tailored to each specific model. A modified recommender system was developed (CaDRReS) in which cell line features were first calculated using gene expression information [94]. The MF method determined the pharmacogenomic space (the dot product of the cell line vector and the drug vector), and drug sensitivity was computed using a specific linear algorithm. The model was compared to other ML algorithms and was found to perform similarly to the elastic net. Because the model provides a projection of cell lines and drugs into the pharmacogenomic space, it is easy to explore relationships between drugs and cell lines [94].

In a recommender system, MF cannot add additional features and cannot predict a completely new item, as the method is highly dependent on data from input features. To resolve those issues, in 2010 Rendle introduced a generalized algorithm, the factorization machine (FM)) [95]. FMs are SVM-like predictors but can handle data with high sparsity (Fig. 8B). Classical FMs can easily handle second-order feature combinations but struggle with higher-order feature combinations. Blondel et al. proposed an updated algorithm for the easy handling of higher-order feature combinations, referred to as higher-order factorization machines (HOFMs) [96]. So far, HOFMs have not been used in monotherapy response prediction; however, they have been employed to predict drug combinations (as described below).

4.4. Autoencoders in monotherapy prediction

An autoencoder is an unsupervised DL model that can be used to reduce the dimension of features. An autoencoder learns hidden (latent) variables from the observed data through the mapping of higher-dimensional data onto a lower-dimensional latent space. An autoencoder consists of two different types of layers: encoding layers and decoding layers, with encoding layers projecting higher-dimensional input data onto lower dimensions and decoding layers reconstructing the lower-dimensional data back to the higher-dimensional data similar to input (Fig. 9A). The loss function is the least squares difference between the input and output vectors. In this case, if the decoding weights correspond to the encoding weights, the output will be the same as the input (deterministic encoding). In general, an autoencoder uses nonlinear activation functions for data compression and can discover nonlinear explanatory features; therefore, it can be used to reduce gene expression features and uncover a biologically relevant latent space [61,97].

Besides the traditional autoencoder, the variational autoencoder (VAE) replaces the deterministic bottleneck layer with stochastic sampling (mean and standard deviation) vectors (Fig. 9B). The model includes regularization losses by adding a Kullback-Leibler (KL) divergence term. This reparameterization allows for backpropagation optimization and for learning the probability distribution of each latent variable instead of directly learning the latent variables [98].

The DL model to predict drug response (DeepDR) combined mutational data with gene expression data to develop a monotherapy prediction model, implementing an autoencoder for both mutational and gene expression data [60]. In this model, the autoencoder was first applied to the TCGA data to transform the mutational and gene expression features into a lower-dimensional representation. The encoded representations of the TCGA data were linked to a feedforward neural network trained on CCLE data for monotherapy prediction. The use of autoencoding
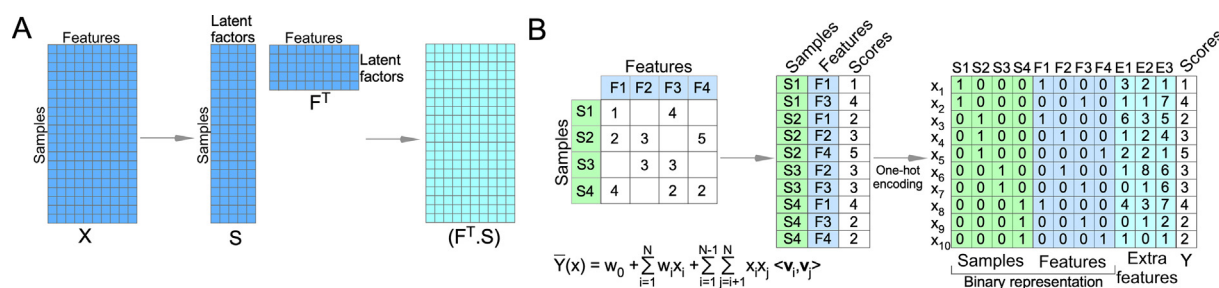

Fig. 8. Matrix factorization and factorization machine. (A) In MF, a matrix is decomposed into two lower-dimensional matrices with the same latent factor. The dot product of lower-dimensional matrices is used to reconstitute the new matrix to calculate the loss function. (B) An FM transforms sample and features data to the binary representation and can incorporate additional features.
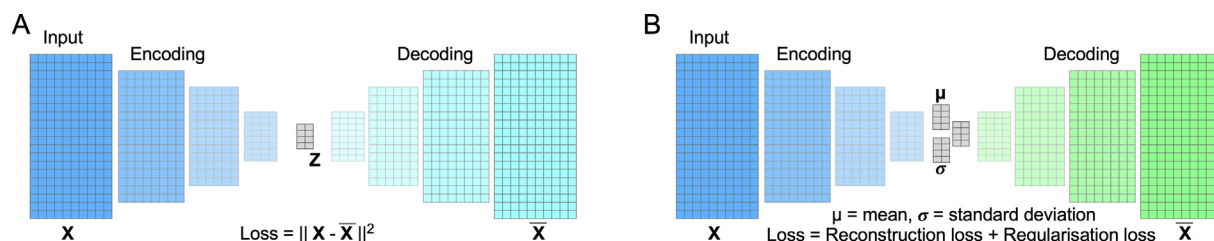
Fig. 9. Autoencoder and variational autoencoder. (A) The autoencoder determines latent variables by reducing the dimensions during encoding. Then it decodes the data into a similar form using the latent variables. (B) VAE uses a similar process unless the latent variables are replaced by the mean and standard deviation.

increased the sample number in the prediction model and, therefore, displayed better prediction performance. Besides an autoencoder, a VAE was used to reduce the higher-dimensional acute myeloid leukemia (AML) patient gene expression data to an 8-dimensional representation, and the VAE was then used to build a linear regression model (lasso) for drug response prediction [99]. Later, a drug response VAE (Dr.VAE) was developed using drug-induced gene expression perturbation [59]. This study used a semi-supervised VAE to predict monotherapy responses using cell line data, and the model was shown to perform better than several linear or nonlinear algorithms. The use of drug-induced gene expression perturbation seems to be useful in determining pathways that regulate drug response and therapy resistance [100]. Nevertheless, anomaly detection with density estimation can improve the prediction accuracy through false positive detection, but this still needs to be implemented [101].

4.5. Graph convolutional networks in monotherapy prediction

Therapy response prediction using multiple drugs requires the incorporation of chemical information about the drugs. This can be done in several ways. The 2D molecular fingerprint (also known as the Morgan fingerprint or circular fingerprint) is commonly measured by the extended-connectivity fingerprint (ECFP) algorithm [102]. This algorithm determines partial structures and converts them into a binary representation. Similarly, the 3D fingerprint descriptor collects 3D information, including electrostatics and molecular shape. The simplified molecular input line entry specification (SMILES) representation was developed by Weininger and provides a linear notation method [103]. SMILES can be used directly by a CNN. Molecular graphs are another type of flexible representation of small-molecule drugs. The GraphDRP study used a molecular graph representation in a GCN to extract molecular features from drugs [104]. At the same time, a CNN was used to extract genomic features from cell lines. Then, the features from the GCN and CNN were combined and fed into the fully connected feedforward neural network for drug sensitivity prediction. The GCN model was compared to a recently developed CNN model using the SMILES format to describe the drugs and was found to perform better, suggesting that the use of graph data for drugs improves predictive performance [105].

4.6. Visible neural networks in monotherapy prediction

Model interpretation is an important research area in ML that seeks to explain the model's internal rationality of a prediction. Biological ML models that were developed with prior knowledge of network or structural data can be explained relatively easily. A so-called visible neural network (VNN) incorporates genomic or transcriptomic data considering the cellular architecture and signaling pathways [62]. Chemical information about drugs was separately processed and then combined with the embedding genotype data to develop the final prediction model (DrugCell).

The DrugCell method was compared to the elastic net and other DNN models and found to have a similar or better predictive performance.

4.7. PDXs and organoids in monotherapy prediction

Although most studies used cell line data to develop ML models, recently the PDXGEM study applied PDXs to develop an ML model [106]. In this study, drug activity was calculated as a percentage of tumor volume changes. Baseline gene expression profiling data were used to develop the model. Another recent study used data from 3D organoid culture models and applied protein–protein interaction networks [107]. The model was trained with pharmacogenomic data from two previous studies using ridge regression [108,109]. This study developed a clinically relevant prediction model that was also useful in identifying predictive biomarkers [107]. Collectively, the use of PDXs and organoids in model development increases the probability of successful clinical applications.

5. Drug synergy prediction

The use of monotherapy in cancer treatment is relatively rare, and most cancer patients are treated with a combination of several drugs. Cancer cells can easily develop resistance to monotherapy, while the development of resistance to several drugs can be difficult or take longer. Therefore, combinatorial therapies are preferred over monotherapy in clinics for cancer treatment. A combination of multiple drugs can have three different effects: additive, antagonistic, and synergistic. The additive effect can be considered a neutral effect, while the antagonistic effect is negative. The synergistic effect is preferable. Thus, predicting drug synergy will be highly beneficial for selecting effective combinations for cancer treatment.

Drug synergy is usually calculated by a cell viability matrix, in which a wide range of single and combinatorial drug effects are noted. The Institute for Molecular Medicine Finland (FIMM) developed an experimental-computational pipeline to measure and visualize synergy from drug combinations [110]. It allows for the simultaneous measurement of several synergy scores, such as Bliss independence [111], Loewe additivity [112], highest single agent (HSA) [113], and zero interaction potency (ZIP) [114]. Later, the study was extended to the prediction of drug combinations [115]. Combenefit is yet another program for calculating synergy scores, in particular Loewe additivity [116].

Several attempts have been made to identify drug synergy using cell lines from different cancers [117–123]. These studies provided an initial framework for developing ML algorithms for predicting drug synergy. A list of available in silico drug synergy prediction models is given in Table 2.

Table 2
Studies predicting drug synergy.

Year	Study name	Data	Algorithm	Ref
2015	RACS	DCDB [151], KEGG, NCI-DREAM	Semi-supervised learning	[118]
2017	Li et al.	DREAM [128]	Random forest	[130]
	Gayvert et al.	Held et al. [120]	Random forest	[140]
	SynGeNet	LINCS L1000, Held et al. [120]	Network-based	[136]
2018	Xia et al.	NCI-ALMANAC [141]	DL	[142]
	Deep Synergy	O'Neil 2016 [122]	DL	[143]
	Deep belief	DREAM [128]	Restricted Boltzmann machine	[150]
2019	SynGeNet	LINCS L1000	Network based	[137]
	DREAM	CNV, mutation, methylation, and gene expression	Multiple	[117]
	DDIGIP	DrugBank, SIDER, OFFSIDES	Regularized Least Squares	[126]
	Cuvitoglu et al.	DCDB [151], Cmap [131]	Naive Bayes, Support Vector Machines, and Random Forest	[132]
	Malyutina et al.	O'Neil 2016 [122]	Elastic net, random forest, support vector machine	[115]
2020	Deep graph	O'Neil 2016 [122]	graph convolutional network	[147]
	comboFM	NCI-ALMANAC [141]	Higher-order factorization machines	[145]
2021	CellBox	Perturbation data [134]	ODE	[133]
	AuDNNsynergy	O'Neil 2016 [122]	Autoencoder	[146]
	TranSynergy	O'Neil 2016 [122]	Transformer boosted DL	[144]

5.1. Drug synergy prediction using conventional ML methods

In silico methods integrating molecular data with pharmacological data could potentially identify drug combinations with some limitations [124]. A heterogeneous network-assisted inference (HNAI) framework was developed using drug-drug interaction pairs connecting approved drugs, phenotypic similarity, therapeutic similarity, chemical structure similarity, and genomic similarity using naive Bayes, decision tree, k-nearest neighbor (KNN), logistic regression, and SVM algorithms [125]. Then, the DDIGIP method, in which the Gaussian interaction profile (GIP) kernel and the regularized least squares (RLS) classifier were implemented, was based on drug-drug interactions (DDIs) [126]. DDIGIP used the similarity of drug features extracted from drug substructures, targets, transporters, enzymes, pathways, indications, side effects, offside effects, and drug-drug interaction data. Collectively, these methods give valuable insights into drug-drug interactions but cannot provide information about whether certain drug combinations will be effective for a specific patient. Gene expression data were used at a limited scale to predict the effect of drug combinations by the Petri net model [127], but the model requires gene expression profiles for every drug pair, which limits its practical applications.

In a DREAM challenge, the human diffuse large B-cell lymphoma (DLBCL) cell line OCI-LY3 was treated with 91 compound pairs of 14 drugs. The drug-induced genomic residual effect model—which combined similarity and dissimilarity in compound activity incorporating drug-induced gene perturbation, dose–response, and pathway information—was reported to outperform 30 other models [128,129]. Although the accuracy of the predictive models was not optimal for practical applications, this study raised the probability of building computational predictive models for drug synergy prediction. The gene expression perturbation data generated in this project are valuable for other studies and can be used to train random forest models with the biological and chemical properties of drugs, such as physicochemical properties, target network distances, and targeted pathways [130]. Similarly, Cuvitoglu et al. extracted the drug perturbation set of genes for each drug from the transcriptome profile of Cmap data [131] and calculated six different features: the distance between two drugs (M1), the mutual information about biological processes (M2), the gene ontology similarity (M3), the overlap of drug perturbation sets (M4), the betweenness centrality of the drug combination network (M5), and the degree of the drug combination network (M6) [132]. Three models were developed using a naive Bayes classifier, an SVM, and a random forest algorithm. Different features were tested, and models combining the M5 and M6 features performed

the best. In addition, the CellBox method used perturbation data of the melanoma SK-Mel-133 cell line treated with 12 different drugs [133,134]. Using nonlinear ordinary differential equations (ODEs), CellBox provided an interpretable ML system that can be used to predict drug combinations in a dynamic system. This study provided mechanistic insights for designing a combination therapy with an understandable predictive model. Taken together, these studies suggest that drug perturbation data provide important information about the regulation of biological features that can be used to develop efficient ML models [100].

Models integrating the signaling network or pathway map have been used to detect drug combinations with limited general applications [135–137]. Similarly, synergy prediction models developed with naive Bayes classifiers [138] and random forest algorithms [139,140] had limited use for specific cell models. Collectively, synergy prediction models developed using classical ML algorithms displayed acceptable predictive performance with specific datasets but largely lacked generalizability.

5.2. Drug synergy prediction using DL

DL has been employed in the prediction of drug synergy. Using the NCI-ALMANAC database [141], it has been demonstrated that the use of gene expression, microRNA, and proteome data, along with drug descriptors, provides the highest prediction capability with feedforward neural networks [142]. This model used two submodels to separately process drug descriptors and gene expression, microRNA, and proteome data. The submodels were fully connected neural networks that helped reduce the dimensionality of the data before they were fed into the final model. This study provided important insight into the use of DL in feature selection and model development.

The DeepSynergy study [143] used a previously published drug synergy dataset [122] to build a DL model and compared it with several classical ML methods, such as gradient boosting, random forest algorithms, SVMs, and elastic nets. This feedforward DL model, which used gene expression data with the chemical features of both drugs to predict Loewe additivity, achieved considerable accuracy. The use of DL allowed the model to perform better than other ML algorithms, but it should also be tested with unknown samples.

Recently, transformer boosted DL (TransSynergy) was developed, in which three components were used: input dimension reduction, a self-attention transformer, and a fully connected output layer [144]. The input vector contained selected features from two drugs (drug-target interaction profile) and the cell line (gene

expression). A fourth dimension was added if both gene expression and gene dependency were used. The use of cell-line-gene dependency, gene-gene interaction, and drug-target interaction provided TransSynergy with a considerably higher predictive performance and allowed the cellular effect of drug actions to be explained. These methods provided a significant improvement over traditional ML mechanisms due to appropriate feature learning. However, all those models used cell line synergy data [122], which might limit their application in preclinical and/or clinical trial settings.

5.3. Synergy prediction with a higher-order factorization machine

An HOFM model [96] was used in comboFM to capture fifth-order feature combinations using data from two drugs, cell lines, and dose–response matrices [145]. The model integrated chemical descriptors of drugs and gene expression data of cell lines as additional features. comboFM was trained with a part of the NCI-ALMANAC data, while the other part of the data was used for predictive performance testing. The fifth-order comboFM was found to perform significantly better than second- and first-order predictors, suggesting that the use of higher-order feature combinations can improve predictive performance.

5.4. Synergy prediction using an autoencoder

An autoencoder has also been employed to predict drug synergy [146]. AuDNNsynergy used multi-omics data from CCLE and TCGA databases combined with previously published drug synergy data [122]. In this study, three independent autoencoders were used to reduce the dimensions of TCGA gene expression, mutation, and copy number data. The reduced dimensions were then combined with drug combination data to develop the model. The model was compared with the recently developed DeepSynergy model and was shown to perform better [143], suggesting that feature reduction using an autoencoder and the use of multi-omics data influence predictive performance.

5.5. Synergy prediction with a graph convolutional network

A graph convolutional network (GCN) model was described (DeepGraph) in which a drug-drug synergy network, a drug-target interaction network, and a protein–protein interaction network were used to build a cell-line-specific model [147]. In the DeepGraph study, a cell-line-specific multirelational network graph was generated and fed into the GCN encoder. A four-layer neural network with a relu activation function was used for encoding, and a sigmoid activation function was used for the embedding output vector. The matrix decoder was used to decode the embedding vector, which predicts the synergy score [74]. The prediction performance of DeepGraph was comparable to that of DeepSynergy. Because the DeepGraph method used a cell-line-specific drug-protein network and protein–protein interaction network and because only limited data for drug-protein interactions were available, the method's performance might be biased.

5.6. Restricted Boltzmann machine for predicting drug synergy

The restricted Boltzmann machine (RBM) is a generative probabilistic model that has been widely used for handling higher-dimensional data [148]. The RBM is similar in function to an autoencoder and can be used to extract meaningful features from higher-dimensional data. Furthermore, multiple RBMs can be stacked to form a deep belief network, which allows unsupervised and supervised data to be combined. RBMs have been used to identify gene expression biomarkers that can help predict clinical out-

comes [149]. Chen et al. used RBMs to develop a deep belief network [150] from the DREAM consortium's drug target information and baseline gene expression data [128]. Although the model was compared with existing DREAM consortium models and was shown to outperform these models, the leave-one-out approach that was adopted in this study was not comparable to the original DREAM consortium models, which were compared with external data.

6. Limitations in the development of clinically relevant predictive models

Currently, most ML models have been developed using cell line data. Cell line data are robust, relatively easy to generate, and useful for hypothesis generation. However, cell line data must be complemented with more disease-relevant patient data. A large-scale pharmacogenomic study using patient data is currently technically difficult because it requires a lot of primary patient materials. This can potentially be overcome by using PDXs. The recent development of PDX repositories will support large-scale clinically relevant studies in the near future [37–40].

Most tumors grow in a multicellular environment in which the surrounding cells create a favorable microenvironment for tumor growth. Prediction models based on cell line data do not capture the microenvironment's contributions and might therefore never reach the level of accuracy that is necessary in the clinic. Cultured tumor organoids can likely mimic the microenvironment of a patient's tumor [107]. However, currently, only limited pharmacogenomic data from tumor organoids are available.

Several recent models used multi-omics data to build predictive models [62,87,92]. Although the use of multi-omics data can improve the prediction performance and can be very useful for research purposes, it limits the practical use of the models in the clinic. For prediction purposes, it would be costly and time-consuming to determine mutations, CNVs, promotor methylation, protein expression, gene expression, etc. for each patient separately. Gene expression data can potentially reflect most cellular processes because mutations, CNVs, and promotor methylation might ultimately determine gene expression changes.

Most gene expression data currently available involve the baseline expression of genes and do not reflect drug-induced perturbations [24,28,30,80]. A few studies provided a limited number of drug-induced perturbation data, which were found to be very useful for feature selection [59,134]. Thus, large-scale drug-induced perturbation studies will help to develop better predictive models.

Nevertheless, drug synergy prediction is an important concept that will have numerous uses in the clinic. At the same time, a combination of several drugs can have severe adverse effects. Thus, a comprehensive method is needed that will not only determine drug synergy but also incorporate the adverse effect of drug combinations. Knowledge of safe and unsafe combinations of drugs was used to build a linear regression prediction model [152–154]. However, the model did not incorporate any biological data to elucidate patient-specific side effects.

Several studies have highlighted implementation challenges encountered in precision medicine solutions [155,156]. These challenges include data preprocessing, unstructured clinical text processing, medical data processing and storage, and environmental data collections. Apart from these challenges, the major challenge might be the redesigning of clinical decision support systems so that they can incorporate molecular, omics, and environmental aspects of precision medicine. A comprehensive support system is desirable to facilitate the curation of data from different sources and multiple scales and to promote the interaction between bioin-

formatics and clinical informatics [155]. Building such a system requires solving many integration and standardization issues.

As pointed out by many studies, model explainability, high-quality training data, and collaborations between medical experts and computational experts are some of the key factors affecting the success of ML solutions for drug response prediction in cancer treatment [9,157]. Although much omics information is available and many theoretical frameworks exist, hands-on ML tools targeted at physicians and medical professionals are scarce. In that regard, various cloud-based cancer prediction tools, such as OASIS-PRO [158], can be introduced to make ML solutions suitable for massive clinical practice. The study gave an overview of general-purpose multi-omics tools that can be useful for gene identification and cancer subtyping [159].

Clinical trials are essential for clinical research in general and cancer treatment in particular. The three-phase trial approach is considered standard practice but is designed primarily for gradually improving treatments. Our ability to understand and treat cancer has, however, evolved over time [21]. Because of the immense role of ML in both clinical trials and clinical practice, the inclusion of ML in regulatory frameworks is unavoidable.

7. Conclusion

The development of predictive models for monotherapy and combinatorial therapies is important but highly challenging. The recent advancement in ML algorithms holds promise for the development of clinically relevant predictive models. Furthermore, more pharmacogenomic data from disease-relevant organoids and PDXs are becoming available, allowing clinical biases to be overcome. Massive computational power is within easy reach for handling a large amount of data that is exponentially increasing. In the near future, the current lack of clinically relevant pharmacogenomic data might also be overcome. Therefore, although current predictive models are far from being ready for clinical use, they show us a clear path toward precision medicine.

CRediT authorship contribution statement

Raihan Rafique: Writing - original draft, Writing - review & editing. **S.M. Riazul Islam:** Writing - original draft, Writing - review & editing. **Julhash U. Kazi:** Conceptualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. 2015;13:8–17.

[2] Sharma A, Rani R. A systematic review of applications of machine learning in cancer prediction and diagnosis. Arch. Comput. Methods Eng. 2021. https://doi.org/10.1007/s11831-021-09556-z.

[3] Hamamoto R, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, et al. Application of artificial intelligence technology in oncology: towards the establishment of precision medicine. Cancers (Basel) 2020;12:3532. https://doi.org/10.3390/cancers12123532.

[4] Putora PM, Baudis M, Beadle BM, El Naqa I, Giordano FA, Nicolay NH. Oncology informatics: status quo and outlook. Oncology 2020;98(Suppl. 6):329–31.

[5] Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer Sci. 2020;111(5):1452–60.

[6] Huang S, Yang J, Fong S, Zhao QI. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. Cancer Lett. 2020;471:61–71.

[7] Nardini C. Machine learning in oncology: a review. Ecancermedicalscience 2020;14:1065.

[8] Filipp FV. Opportunities for artificial intelligence in advancing precision medicine. Curr. Genet. Med. Rep. 2019;7(4):208–13.

[9] Azuaje F. Artificial intelligence for precision oncology: beyond patient stratification. NPJ Precis. Oncol. 2019;3:6.

[10] Patel SK, George B, Rai V. Artificial intelligence to decode cancer mechanism: beyond patient stratification for precision oncology. Front. Pharmacol. 2020;11:1177.

[11] J. Singer, A. Irmisch, H.J. Ruscheweyh, F. Singer, N.C. Toussaint, M.P. Levesque, D.J. Stekhoven, N. Beerenwinkel, Bioinformatics for precision oncology. Brief Bioinform 20 (2019) 778–788.

[12] Nicora G, Vitali F, Dagliati A, Geifman N, Bellazzi R. Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. Front. Oncol. 2020;10:1030.

[13] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat. Rev. Clin. Oncol. 2019;16(11):703–15.

[14] Grothen AE, Tennant B, Wang C, Torres A, Bloodgood Sheppard B, Abastillas G, et al. Application of artificial intelligence methods to pharmacy data for cancer surveillance and epidemiology research: a systematic review. JCO Clin. Cancer Inform. 2020(4):1051–8. https://doi.org/10.1200/CCI.20.00101.

[15] Biswas N, Chakrabarti S. Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. Front. Oncol. 2020;10:588221.

[16] Chiu YC, Chen HH, Gorthi A, Mostavi M, Zheng S, Huang Y, et al. Deep learning of pharmacogenomics resources: moving towards precision oncology. Brief Bioinform 2020;21:2066–83. https://doi.org/10.1093/bib/bbz144.

[17] Adam G, Rampasek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. NPJ Precis. Oncol. 2020;4:19.

[18] Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in oncology: a clinical appraisal. Cancer Lett. 2020;481:55–62.

[19] Tanoli Z, Vaha-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. Expert Opin. Drug Discov. 2021:1–13.

[20] Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clin. Epigenet. 2020;12:51.

[21] Li A, Bergan RC. Clinical trial design: past, present, and future in the context of big data and precision medicine. Cancer 2020;126(22):4838–46.

[22] Fountzilas E, Tsimberidou AM. Overview of precision oncology trials: challenges and opportunities. Expert. Rev. Clin. Pharmacol. 2018;11 (8):797–804.

[23] Li X, Warner JL. A review of precision oncology knowledgebases for determining the clinical actionability of genetic variants. Front. Cell Dev. Biol. 2020;8:48.

[24] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483(7391):603–7.

[25] Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature 2019;569(7757):503–8.

[26] Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer 2006;6(10):813–23.

[27] Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. Cell 2016;166 (3):740–54.

[28] Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 2012;483(7391):570–5.

[29] Haverty PM, Lin E, Tan J, Yu Y, Lam B, Lianoglou S, et al. Reproducible pharmacogenomic profiling of cancer cell line panels. Nature 2016;533 (7603):333–7.

[30] Basu A, Bodycombe N, Cheah J, Price E, Liu Ke, Schaefer G, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 2013;154(5):1151–61.

[31] Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. 2015;5(11):1210–23.

[32] Smirnov P, Kofia V, Maru A, Freeman M, Ho C, El-Hachem N, et al. PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. Nucleic Acids Res 2018;46:D994–D1002. https://doi.org/10.1093/nar/gkx911.

[33] Rajapakse VN, Luna A, Yamade M, Loman L, Varma S, Sunshine M, et al. CellMinerCDB for integrative cross-database genomics and

pharmacogenomics analyses of cancer cell lines. iScience 2018;10:247–64. https://doi.org/10.1016/j.isci.2018.11.029.

[34] Luna A, Elloumi F, Varma S, Wang Y, Rajapakse VN, Aladjem MI, et al. Cell Miner Cross-Database (CellMinerCDB) version 1.2: Exploration of patient-derived cancer cell line pharmacogenomics. Nucleic Acids Res. 2021;49: D1083–93.

[35] Zagidullin B, Aldahdooh J, Zheng S, Wang W, Wang Y, Saad J, et al. DrugComb: an integrative cancer drug combination data portal. Nucleic Acids Res 2019;47:W43–51. https://doi.org/10.1093/nar/gkz337.

[36] Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Res. 2020;48:D871–81.

[37] Conte N, Mason JC, Halmagyi C, Neuhauser S, Mosaku A, Yordanova G, et al. PDX Finder: A portal for patient-derived tumor xenograft model discovery. Nucleic Acids Res 2019;47:D1073–9. https://doi.org/10.1093/nar/gky984.

[38] Townsend EC, Murakami MA, Christodoulou A, Christie AL, Köster J, DeSouza TA, et al. The Public Repository of Xenografts Enables Discovery and Randomized Phase II-like Trials in Mice. Cancer Cell 2016;29(4):574–86.

[39] PDMR (2021) NCI's Patient-derived Models Repository. https://pdmr.cancer.gov/.

[40] Hidalgo M, Amant F, Biankin AV, Budinská E, Byrne AT, Caldas C, et al. Patient-derived xenograft models: an emerging platform for translational cancer research. Cancer Discov. 2014;4(9):998–1013.

[41] Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat. Med. 2015;21(11):1318–25.

[42] Mer AS, Ba-Alawi W, Smirnov P, Wang YX, Brew B, Ortmann J, et al. Integrative Pharmacogenomics Analysis of Patient-Derived Xenografts. Cancer Res. 2019;79(17):4539–50.

[43] Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat. Biotechnol. 2015;33(3):306–12.

[44] Greshock J, Bachman KE, Degenhardt YY, Jing J, Wen YH, Eastman S, et al. Molecular target class is predictive of in vitro response profile. Cancer Res. 2010;70(9):3677–86.

[45] Mpindi JP, Yadav B, Östling P, Gautam P, Malani D, Murumägi A, et al. Consistency in drug response profiling. Nature 2016;540(7631):E5–6.

[46] Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJWL, et al. Inconsistency in large pharmacogenomic studies. Nature 2013;504 (7480):389–93.

[47] The Cancer Cell Line Encyclopedia and Genomics of Drug Sensitivity in Cancer Investigators. Pharmacogenomic agreement between two cancer cell line data sets. Nature 2015;528:84–7. https://doi.org/10.1038/nature15736.

[48] Safikhani Z, Smirnov P, Thu KL, Silvester J, El-Hachem N, Quevedo R, et al. Gene isoforms as expression-based biomarkers predictive of drug response in vitro. Nat. Commun. 2017;8(1). https://doi.org/10.1038/s41467-017-01153-8.

[49] Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. 2014;32(12):1202–12.

[50] Jang IS, Neto EC, Guinney J, Friend SH, Margolin AA. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. Pac. Symp. Biocomput. 2014:63–74.

[51] Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys. Rev. 2019;11(1):31–9. https://doi.org/10.1007/s12551-018-0446-z.

[52] Koras K, Juraeva D, Kreis J, Mazur J, Staub E, Szczurek E. Feature selection strategies for drug sensitivity prediction. Sci. Rep. 2020;10:9377.

[53] Ali M, Khan SA, Wennerberg K, Aittokallio T. Global proteomics profiling improves drug sensitivity prediction: results from a multi-omics, pan-cancer modeling approach. Bioinformatics 2018;34:1353–62. https://doi.org/10.1093/bioinformatics/btx766.

[54] Papillon-Cavanagh S, De Jay N, Hachem N, Olsen C, Bontempi G, Aerts HJWL, et al. Comparison and validation of genomic predictors for anticancer drug sensitivity. J. Am. Med. Inform. Assoc. 2013;20(4):597–602.

[55] Stetson LC, Pearl T, Chen Y, Barnholtz-Sloan JS. Computational identification of multi-omic correlates of anticancer therapeutic response. BMC Genom. 2014;15(Suppl 7):S2. https://doi.org/10.1186/1471-2164-15-S7-S2.

[56] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 2005;03(02):185–205.

[57] Lin TH, Li HT, Tsai KC. Implementing the Fisher's discriminant ratio in a k-means clustering algorithm for feature selection and data set trimming. J. Chem. Inf. Comput. Sci. 2004;44(1):76–87.

[58] Nakajo M, Jinguji M, Tani A, Hirahara D, Nagano H, Takumi K, et al. Application of a machine learning approach to characterization of liver function using (99m)Tc-GSA SPECT/CT. Abdom Radiol (NY) 2021;46 (7):3184–92. https://doi.org/10.1007/s00261-021-02985-1.

[59] Rampasek L, Hidru D, Smirnov P, Haibe-Kains B, Goldenberg A. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. Bioinformatics 2019;35:3743–51.

[60] Chiu YC, Chen HIH, Zhang T, Zhang S, Gorthi A, Wang LJ, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC Med. Genom. 2019;12(S1). https://doi.org/10.1186/s12920-018-0460-9.

[61] Li M, Wang Y, Zheng R, Shi X, Li Y, Wu FX, et al. DeepDSC: a deep learning method to predict drug sensitivity of cancer cell lines. IEEE/ACM Trans.

Comput. Biol. Bioinform. 2021;18(2):575–82. https://doi.org/10.1109/TCBB.2019.2919581.

[62] Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell 2020;38(672–684):e676.

[63] Liu X, Li N, Liu S, Wang J, Zhang N, Zheng X, et al. Normalization methods for the analysis of unbalanced transcriptome data: a review. Front. Bioeng. Biotechnol. 2019;7. https://doi.org/10.3389/fbioe.2019.00358.

[64] Reinders C, Ackermann H, Yang MY, Rosenhahn B. Learning convolutional neural networks for object detection with very little training data. Multimodal Scene Understanding 2019:65–100.

[65] Jonsson A. Deep reinforcement learning in medicine. Kidney Dis. (Basel) 2019;5(1):18–22.

[66] Triguero I, García S, Herrera F. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowl. Inf. Syst. 2015;42 (2):245–84.

[67] Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. J. Med. Syst. 2002;26:445–63.

[68] Breiman L. Random forests. Mach. Learn. 2001;45:5–32.

[69] Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. Computational Learning Theory, EuroCOLT 1995;904:23–37.

[70] Friedman J, Hastie T, Tibshirani R. Special Invited Paper. Additive logistic regression: a statistical view of boosting. Ann. Stat. 2000;28:337–74.

[71] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. arXiv 2016, 1603.02754v3. https://doi.org/10.1145/2939672.2939785.

[72] Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw. 2001;12(2):181–201.

[73] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.

[74] Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv 2017:1609.02907v4.

[75] Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F. Graph convolutional networks for computational drug development and discovery. Brief Bioinform 2020;21:919–35.

[76] van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530–6.

[77] Kaklamani V. A genetic signature can predict prognosis and response to therapy in breast cancer: oncotype DX. Expert Rev. Mol. Diagn. 2006;6 (6):803–9.

[78] Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. 2009;27(8):1160–7.

[79] Geeleher P, Cox NJ, Huang R. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol. 2014;15(3):R47. https://doi.org/10.1186/gb-2014-15-3-r47.

[80] Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS ONE 2013;8(4):e61318. https://doi.org/10.1371/journal.pone.0061318.

[81] Cortes-Ciriano I, van Westen GJ, Bouvier G, Nilges M, Overington JP, Bender A, et al. Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. Bioinformatics 2016;32:85–95.

[82] Zhang N, Wang H, Fang Y, Wang J, Zheng X, Liu XS, et al. Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. PLoS Comput. Biol. 2015;11(9):e1004498. https://doi.org/10.1371/journal.pcbi.1004498.

[83] Zhang F, Wang M, Xi J, Yang J, Li A. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. Sci. Rep. 2018;8:3355.

[84] Gonen M, Margolin AA. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. Bioinformatics 2014;30: i556–563.

[85] Ammad-Ud-Din M, Khan SA, Wennerberg K, Aittokallio T. Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. Bioinformatics 2017;33: i359–68.

[86] Ammad-ud-din M, Georgii E, Gönen M, Laitinen T, Kallioniemi O, Wennerberg K, et al. Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization. J. Chem. Inf. Model. 2014;54(8):2347–59.

[87] Lee SI, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat. Commun. 2018;9(1). https://doi.org/10.1038/s41467-017-02465-5.

[88] Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. Mol. Pharm. 2016;13 (7):2524–30.

[89] Deng L, Cai Y, Zhang W, Yang W, Gao B, Liu H. Pathway-Guided Deep Neural Network toward Interpretable and Predictive Modeling of Drug Sensitivity. J. Chem. Inf. Model. 2020;60(10):4497–505.

[90] Sakellaropoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, et al. A Deep Learning Framework for Predicting Response to Therapy in Cancer. Cell Rep 2019;29(11):3367–3373.e4.

[91] Chang Y, Park H, Yang HJ, Lee S, Lee KY, Kim TS, et al. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug

Effectiveness from Cancer Genomic Signature. Sci. Rep. 2018;8(1). https://doi.org/10.1038/s41598-018-27214-6.

[92] Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester MMOLI. multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics 2019;35:i501–9.

[93] Koren Y, Bell R, Volinsky C. Matrix Factorization Technique for Recommender System. Computer 2009;42:30–7.

[94] Suphavilai C, Bertrand DN. Nagarajan. Predicting Cancer Drug Response using a Recommender System. Bioinformatics 2018;34:3907–14.

[95] Rendle S. Factorization Machines. IEEE International Conference on Data Mining IEEE 2010;2010:995–1000.

[96] Blondel M, Fujino A, Ueda N, Ishihata M. Higher-Order Factorization Machines. In: 30th Conference on Neural Information Processing Systems NIPS. p. 3351–9.

[97] Way GP, Greene CS. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. Pac Symp Biocomput 2018;23:80–91.

[98] Kingma DP, Welling M. Auto-Encoding Variational Bayes. arXiv 2014:1312.6114.

[99] Dincer AV, Celik S, Hiranuma N, LeeDeepProfile SI. Deep learning of cancer molecular profiles for precision medicine. bioRxiv 2018. https://doi.org/10.1101/278739.

[100] Shah K, Ahmed M, Kazi JU. The Aurora kinase/beta-catenin axis contributes to dexamethasone resistance in leukemia. npj Precis. Oncol. 2021;5:13. https://doi.org/10.1038/s41698-021-00148-5.

[101] Nachman B, Shih D. Anomaly detection with density estimation. Phys Rev D 2020;101(7). https://doi.org/10.1103/PhysRevD.101.075042.

[102] Rogers D, Hahn M. Extended-connectivity fingerprints. J. Chem. Inf. Model. 2010;50(5):742–54.

[103] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 1988;28:31–6.

[104] Nguyen TT, Nguyen GTT, Nguyen T, Le DH. Graph convolutional networks for drug response prediction. IEEE/ACM Trans Comput Biol Bioinform 2021. https://doi.org/10.1109/TCBB.2021.3060430.

[105] Liu P, Li H, Li S, Leung KS. Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. BMC Bioinf. 2019;20:408.

[106] Kim Y, Kim D, Cao B, Carvajal R, Kim M. PDXGEM: patient-derived tumor xenograft-based gene expression model for predicting clinical response to anticancer therapy in cancer patients. BMC Bioinf. 2020;21:288.

[107] Kong JH, Lee H, Kim D, Han SK, Ha D, Shin K, et al. Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. Nat. Commun. 2020;11(1). https://doi.org/10.1038/s41467-020-19313-8.

[108] van de Wetering M, Francies HE, Francis JM, Bounova G, Iorio F, Pronk A, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell 2015;161(4):933–45.

[109] Lee SH, Hu W, Matulay JT, Silva MV, Owczarek TB, Kim K, et al. Tumor Evolution and Drug Response in Patient-Derived Organoid Models of Bladder Cancer. Cell 2018;173(2):515–528.e17.

[110] He L, Kulesskiy E, Saarela J, Turunen L, Wennerberg K, Aittokallio T, et al. Methods for High-throughput Drug Combination Screening and Synergy Scoring. Methods Mol. Biol. 2018;1711:351–98.

[111] Bliss CI. The toxicity of poisons applied jointly. Ann. Appl. Biol. 1939;26:585–615.

[112] Loewe S. The problem of synergism and antagonism of combined drugs. Arzneimittelforschung 1953;3:285–90.

[113] Tan X, Hu L, Luquette LJ, Gao G, Liu Y, Qu H, et al. Systematic identification of synergistic drug pairs targeting HIV. Nat. Biotechnol. 2012;30(11):1125–30.

[114] Yadav B, Wennerberg K, Aittokallio T, Tang J. Searching for Drug Synergy in Complex Dose-Response Landscapes Using an Interaction Potency Model. Comput. Struct. Biotechnol. J. 2015;13:504–13.

[115] Malyutina A, Majumder MM, Wang W, Pessia A, Heckman CA, Tang J, et al. Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. PLoS Comput. Biol. 2019;15(5): e1006752. https://doi.org/10.1371/journal.pcbi.1006752.

[116] Di Veroli GY, Fornari C, Wang D, Mollard S, Bramhall JL, Richards FM, et al. Combenefit: an interactive platform for the analysis and visualization of drug combinations. Bioinformatics 2016;32(18):2866–8.

[117] Menden MP, Wang D, Mason MJ, Szalai B, Bulusu KAC, Guan Y, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat. Commun. 2019;10(1). https://doi.org/10.1038/s41467-019-09799-2.

[118] Sun Y, Sheng Z, Ma C, Tang K, Zhu R, Wu Z, et al. Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. Nat. Commun. 2015;6(1). https://doi.org/10.1038/ncomms9481.

[119] Roller DG, Axelrod M, Capaldo BJ, Jensen K, Mackey A, Weber MJ, et al. Synthetic lethal screening with small-molecule inhibitors provides a pathway to rational combination therapies for melanoma. Mol. Cancer Ther. 2012;11(11):2505–15.

[120] Held MA, Langdon CG, Platt JT, Graham-Steed T, Liu Z, Chakraborty A, et al. Genotype-selective combination therapies for melanoma identified by high-throughput drug screening. Cancer Discov 2013;3(1):52–67.

[121] Kang Y, Hodges A, Ong E, Roberts W, Piermarocchi C, Paternostro G, et al. Identification of drug combinations containing imatinib for treatment of BCR-ABL+ leukemias. PLoS ONE 2014;9(7):e102221. https://doi.org/10.1371/journal.pone.0102221.

[122] O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, et al. An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. Mol. Cancer Ther. 2016;15(6):1155–62.

[123] Chan GKY, Wilson S, Schmidt S, Moffat JG. Unlocking the Potential of High-Throughput Drug Combination Assays Using Acoustic Dispensing. J Lab Autom 2016;21(1):125–32.

[124] Zhao XM, Iskar M, Zeller G, Kuhn M, van Noort V, Bork P, et al. Prediction of drug combinations by integrating molecular and pharmacological data. PLoS Comput. Biol. 2011;7(12):e1002323. https://doi.org/10.1371/journal.pcbi.1002323.

[125] Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inform Assoc 2014;21:e278–286.

[126] Yan C, Duan G, Pan Y, Wu FX, Wang J. DDIGIP: predicting drug-drug interactions based on Gaussian interaction profile kernels. BMC Bioinf. 2019;20:538.

[127] Jin G, Zhao H, Zhou X, Wong STC. An enhanced Petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data. Bioinformatics 2011;27(13):i310–6.

[128] Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, et al. A community computational challenge to predict the activity of pairs of compounds. Nat. Biotechnol. 2014;32(12):1213–22.

[129] Goswami CP, Cheng L, Alexander PS, Singal A, Li L. A New Drug Combinatory Effect Prediction Algorithm on the Cancer Cell Based on Gene Expression and Dose-Response Curve. CPT Pharmacometrics Syst Pharmacol 2015;4 (2):80–90.

[130] Li X, Xu Y, Cui H, Huang T, Wang D, Lian B, 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.

[131] Lamb J. The Connectivity Map: a new tool for biomedical research. Nat. Rev. Cancer 2007;7(1):54–60.

[132] Cuvitoglu A, Zhou JX, Huang S, Isik Z. Predicting drug synergy for precision medicine using network biology and machine learning. J Bioinform Comput Biol 2019;17(02):1950012. https://doi.org/10.1142/S0219720019500124.

[133] Yuan B, Shen C, Luna A, Korkut A, Marks DS, Ingraham J, et al. Cell Box: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. Cell Syst 2021;12(128–140): e124.

[134] Korkut A, Wang W, Demir E, Aksoy BA, Jing X, Molinelli EJ, et al. Perturbation biology nominates upstream-downstream drug combinations in RAF inhibitor resistant melanoma cells. Elife 2015;4:e04640. https://doi.org/10.7554/eLife.04640.

[135] Morris MK, Clarke DC, Osimiri LC, Lauffenburger DA. Systematic Analysis of Quantitative Logic Model Ensembles Predicts Drug Combination Effects on Cell Signaling Networks. CPT Pharmacometrics Syst Pharmacol 2016;5 (10):544–53.

[136] Regan KE, Payne PRO, Li F. Integrative network and transcriptomics-based approach predicts genotype- specific drug combinations for melanoma. AMIA Jt Summits Transl Sci Proc 2017;2017:247–56.

[137] Regan-Fendt KE, Xu J, DiVincenzo M, Duggan MC, Shakya R, Na R, et al. Synergy from gene expression and network mining (SynGeNet) method predicts synergistic drug combinations for diverse melanoma genomic subtypes. npj Syst. Biol. Appl. 2019;5(1). https://doi.org/10.1038/s41540-019-0085-4.

[138] Li P, Huang C, Fu Y, Wang J, Wu Z, Ru J, et al. Large-scale exploration and analysis of drug combinations. Bioinformatics 2015;31(12):2007–16.

[139] Wildenhain J, Spitzer M, Dolma S, Jarvik N, White R, Roy M, et al. Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. Cell Syst 2015;1(6):383–95.

[140] Gayvert KM, Aly O, Platt J, Bosenberg MW, Stern DF, Elemento O, et al. A Computational Approach for Identifying Synergistic Drug Combinations. PLoS Comput. Biol. 2017;13(1):e1005308. https://doi.org/10.1371/journal.pcbi.1005308.

[141] Holbeck SL, Camalier R, Crowell JA, Govindharajulu JP, Hollingshead M, Anderson LW, 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(13):3564–76.

[142] Xia F, Shukla M, Brettin T, Garcia-Cardona C, Cohn J, Allen JE, et al. Predicting tumor cell line response to drug pairs with deep learning. BMC Bioinf. 2018;19(S18). https://doi.org/10.1186/s12859-018-2509-3.

[143] Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. Bioinformatics 2018;34:1538–46. https://doi.org/10.1093/bioinformatics/btx806.

[144] Liu Q, Xie L. TranSynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. PLoS Comput. Biol. 2021;17:e1008653.

[145] Julkunen H, Cichonska A, Gautam P, Szedmak S, Douat J, Pahikkala T, et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. Nat. Commun. 2020;11(1). https://doi.org/10.1038/s41467-020-19950-z.

[146] Zhang T, Zhang L, Payne PRO, Li F. Synergistic Drug Combination Prediction by Integrating Multiomics Data in Deep Learning Models. Methods Mol. Biol. 2021;2194:223–38.

[147] Jiang P, Huang S, Fu Z, Sun Z, Lakowski TM, Hu P. Deep graph embedding for prioritizing synergistic anticancer drug combinations. Comput. Struct. Biotechnol. J. 2020;18:427–38.

[148] Larochelle H, Mandel M, Pascanu Y, Bengio Y. Learning algorithms for the classification restricted boltzmann machine. J. Mach. Learn. Res. 2012;12:643–69.

[149] Jin T, Nguyen ND, Talos F, Wang D. ECMarker: interpretable machine learning model identifies gene expression biomarkers predicting clinical outcomes and reveals molecular mechanisms of human disease in early stages. Bioinformatics 2020.

[150] Chen G, Tsoi A, Xu H, Zheng WJ. Predict effective drug combination by deep belief network and ontology fingerprints. J. Biomed. Inform. 2018;85:149–54.

[151] Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X. DCDB 2.0: a major update of the drug combination database. Database (Oxford) 2014;124. https://doi.org/10.1093/database/bau124.

[152] Huang H, Zhang P, Qu XA, Sanseau P, Yang L. Systematic prediction of drug combinations based on clinical side-effects. Sci. Rep. 2014;4:7160.

[153] Torres NB, Altafini C. Drug combinatorics and side effect estimation on the signed human drug-target network. BMC Syst. Biol. 2016;10:74.

[154] Gardiner LJ, Carrieri AP, Wilshaw J, Checkley S, Pyzer-Knapp EO, Krishna R. Using human in vitro transcriptome analysis to build trustworthy machine learning models for prediction of animal drug toxicity. Sci. Rep. 2020;10:9522.

[155] Afzal M, Islam SMR, Hussain M, Lee S. Precision Medicine Informatics: Principles, Prospects, and Challenges. IEEE Access 2020;8:13593–612.

[156] Kaur S, Singla J, Nkenyereye L, Jha S, Prashar D, Joshi GP, et al. Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. IEEE Access 2020;8:228049–69.

[157] Zhu W, Xie L, Han J, Guo X. The Application of Deep Learning in Cancer Prognosis Prediction. Cancers (Basel) 2020:12. https://doi.org/10.3390/cancers12030603.

[158] Yu KH, Fitzpatrick MR, Pappas L, Chan W, Kung J, Snyder M. Omics AnalySIs System for PRecision Oncology (OASISPRO): a web-based omics analysis tool for clinical phenotype prediction. Bioinformatics 2018;34:319–20. https://doi.org/10.1093/bioinformatics/btx572.

[159] Sathyanarayanan A, Gupta R, Thompson EW, Nyholt DR, Bauer DC, Nagaraj SH. A comparative study of multi-omics integration tools for cancer driver gene identification and tumour subtyping. Brief Bioinform 2020;21:1920–36. https://doi.org/10.1093/bib/bbz121.