ANNUAL REVIEWS

Annual Review of Pharmacology and Toxicology

Artificial Intelligence in Drug Treatment

Eden L. Romm[1] and Igor F. Tsigelny[1,2]

[1]CureMatch Inc., San Diego, California 92121, USA

[2]San Diego Supercomputer Center, University of California, San Diego, La Jolla, California 92093, USA; email: itsigeln@ucsd.edu

Annu. Rev. Pharmacol. Toxicol. 2020. 60:353–69

First published as a Review in Advance on July 26, 2019

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

https://doi.org/10.1146/annurev-pharmtox-010919-023746

Keywords

artificial intelligence, machine learning, deep learning, personalized medicine, drug combination, combination therapy

Abstract

The most common applications of artificial intelligence (AI) in drug treatment have to do with matching patients to their optimal drug or combination of drugs, predicting drug-target or drug-drug interactions, and optimizing treatment protocols. This review outlines some of the recently developed AI methods aiding the drug treatment and administration process. Selection of the best drug(s) for a patient typically requires the integration of patient data, such as genetics or proteomics, with drug data, like compound chemical descriptors, to score the therapeutic efficacy of drugs. The prediction of drug interactions often relies on similarity metrics, assuming that drugs with similar structures or targets will have comparable behavior or may interfere with each other. Optimizing the dosage schedule for administration of drugs is performed using mathematical models to interpret pharmacokinetic and pharmacodynamic data. The recently developed and powerful models for each of these tasks are addressed, explained, and analyzed here.

INTRODUCTION

The era of personalized medicine has begun. Recent advancements in the genomic space have provided a great amount of genomic and transcriptomic data relating to diseases, healthy tissues, animals, and cell lines (1). Thousands of quantitative structure–activity relationship (QSAR) descriptors and molecular fingerprints have been developed for the quantitative and categorical characterization of drug molecules. Computational algorithms, especially related to artificial intelligence (AI), and parallel processing have seen significant advancement in recent decades, giving computer-based inference engines the ability to reach ever-deeper conclusions. This has culminated in the compilation of many databases that are rich in biomedical information and the creation of a variety of computational tools and AI-based software packages for researchers to incorporate into their work. Researchers have leveraged these newly available computational methods and data in developing engines to aid in the prediction of patient-specific drug treatments and optimization of dosage and drug schedules to ensure the highest likelihood of success and quality of life for patients. In this review, we discuss AI-based, single-drug prediction engines and drug-target interactions. We then discuss drug combination therapy (CT), including synergism predictions and drug-drug interactions (DDIs). Lastly, we address AI systems for dosage control and drug regimen optimization.

MONOTHERAPIES: SINGLE-DRUG DECISION SUPPORT SYSTEMS

Prediction of the ideal drug for a given patient's disease can have broad treatment strategies that take into consideration all the relevant medical information from a patient, including genomics and/or proteomics data, integrated with features describing the drug's properties. Often, it is necessary to predict the drug's interactions with biomolecules such as proteins, DNA, RNA, microRNA, and others. The clinical application of such systems will necessitate a combination of these approaches, especially for complex and orphan diseases in which novel cases are common.

Menden and colleagues (2) developed a model for predicting IC_{50} values by integrating the mutation profiles of 77 genes, microsatellite instability, and copy number alterations of 608 cell lines with the QSAR descriptors of 108 drugs. QSAR descriptors numerically represent the various physiochemical properties of small molecules. The training data set contained 38,930 drug–cell line pairs with known IC_{50} values, and the testing set contained 13,565 such pairs (2). They were able to achieve an R^2 value of 0.72 between actual and predicted IC_{50} values on the training set and an R^2 of 0.64 for the testing set, using both neural network and random forest (RF) architectures in separate models. [R^2 is a scalar measure, between 0 and 1, of how well a series of predicted values match the expected values, with a value of 1 indicating perfect prediction and 0 indicating no correlation between calculated outputs and real ones. An R^2 value $\geq$0.7 is considered acceptable for validation on the training set, and an R^2 value $\geq$0.6 on the testing set validates the model as predictive (3).] The Menden et al. model (2) meets both abovementioned criteria, with 79% concordance between genetic relationships determined from predicted values and experimental values, providing a strong case for the dependence of drug resistance on these genes (2). The confirmation of computationally derived gene roles by in vitro experimentation supports the applicability of such models beyond the research setting and is an important step toward the adoption of such systems.

Chang and colleagues (4) developed a more accurate model using an updated version of the same data source (Genomics of Drug Sensitivity in Cancer, or GDSC; **https://www.cancerrxgene.org**) for cell line–drug pairs. Two interesting improvements to the modeling approach were implemented. First, they utilized deep learning (DL) (4); DL is a machine learning (ML) approach dependent on many hidden layers for modeling complex, nonlinear systems (5).

Second, they employed a committee approach, building five DL models with various parameters for the prediction of IC_{50} values and taking the mean of their outputs as the final prediction. The determination of IC_{50} values by taking the mean of five models' outputs (4) achieved an increase of 0.12 in R^2, or a total R^2 value of 0.84, and an area under the receiver operating characteristic curve (AUC) of 0.98 on the training set. [AUC is another common scalar measure of a model's predictive power. AUC values over 0.9, as in Chang et al.'s model (4), are indicative of excellent predictive performance, and any value over 0.7 is considered acceptable.] Both models (2, 4) consider all interactions leading to a single value of drug activity.

Prediction of drug-target interactions is a smaller part of the monotherapy optimization problem, but it is no less important. Predicting drug-target interactions is vital to the repurposing of already existing drugs and the search for novel drugs, along with understanding the signaling and metabolic pathways involved. Attempts at predicting new uses for existing drugs often rely on the use of genetic, transcriptomic, and side effect data interpreted through an AI engine.

The predictions of drug-target interactions presented here rely mainly on two branches of AI techniques: network theory approaches, in which the input characterizes the protein networks with which a drug interacts (6–8), and ML approaches, which utilize architectures like Support Vector Machine (SVM) (9). Campillos et al. (9) constructed a model relying on drug chemical structure and therapeutic similarity indices to predict new and unexpected drug-target interactions using an SVM architecture. [SVM is an ML method that aims to separate categorical variables by drawing multidimensional surfaces between them (**Figure 1**).] The authors described unexpected drug-target interactions for drugs that therapeutically have similar side effects but are not structurally similar (9). For example, cetirizine, which targets HRH1, was found to be therapeutically similar to acitretin, a molecule without previously known interactions with HRH1 but that is structurally

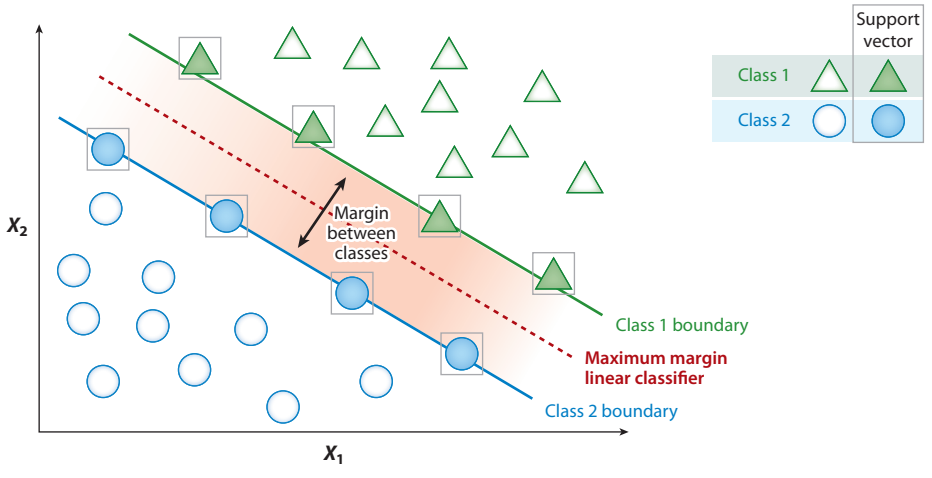


Figure 1

Diagram of a simplified Support Vector Machine architecture for a two-category classification problem. Category boundaries are represented by solid lines and are separated by the maximum margin linear classifier (*dashed line*), representative of the multidimensional surface referenced in the text. One could imagine the two categories (*green triangles* and *blue circles*) here to be active and inactive molecules for a specific protein. The number of molecules placed on the correct side of the line determines the accuracy of the model. In reality, this image can be much more complicated, with no single, straight line being able to achieve complete separation between the categories. Figure adapted from Reference 10.

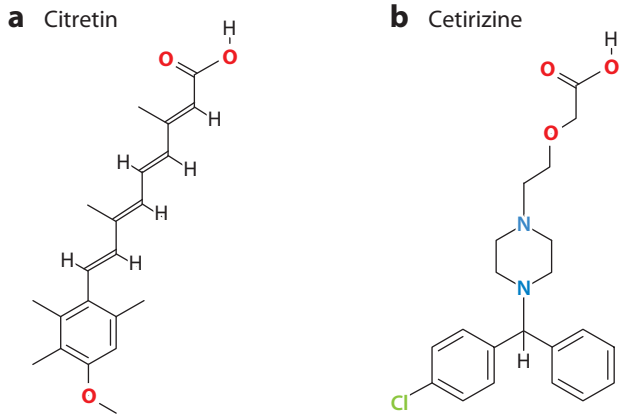
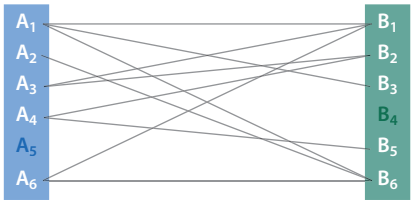


Figure 2

2-D chemical structures of the drugs (*a*) acitretin (60) and (*b*) cetirizine (61), illustrating how different the structures of these two drugs are despite their observed therapeutic similarity. The observed differences are represented by a chemical similarity metric, the 2-D Tanimoto coefficient, between 0.22 and 0.26, indicative of significant structural differences (9). Figure adapted from References 60 and 61.

quite different (**Figure 2**). Acitretin's novel activity on HRH1 was then confirmed experimentally in an assay demonstrating a K_i of 15 μM (9). Campillos and coauthors (9) were able to predict 754 previously unknown interactions; they tested 20 of these in vitro, finding 13 of them to be active on at least one of their predicted targets and 11 of them to have K_i values below 10 μM. Cheng and coauthors (6) built similar models, instead utilizing a number of similarity-based modeling mechanisms; drug-based similarity inference; target-based similarity inference; and a bipartite graph network-based approach, termed network-based inference (NBI), to predict novel drug-target interactions. [A bipartite graph is a representation method in which vertices, illustrating interactions, are drawn only between disjoint sets (**Figure 3**).] NBI proved to be the most powerful method and was used to test further data. Cheng and colleagues (6) applied the NBI model to predict novel

a



b

	B_1	B_2	B_3	B_4	B_5	B_6
A_1	1	0	1	0	0	1
A_2	0	0	0	0	0	1
A_3	1	1	0	0	0	0
A_4	0	1	0	0	1	0
A_5	0	0	0	0	0	0
A_6	1	0	0	0	0	1

0 = Interaction absent 1 = Interaction present

Figure 3

(*a*) Bipartite graph where connections (*gray lines*) between objects in disjoint sets A_i and B_i represent interactions. The disjoint sets could be representative of proteins that may interact with those in the other set but not within the same set. A given protein in the set A_i will not interact with another protein in the set A_i but may interact with any or all proteins in set B_i. (*b*) Adjacency matrix from which the bipartite graph in panel *a* was made. The presence of 1 indicates an interaction between the two elements intersecting at that point, and the presence of 0 illustrates the absence of an interaction between the two elements. Notice that A_5 and B_4 appear in the adjacency matrix but not the bipartite graph in panel *a*, as they do not have any interactions at all, shown by the zeros in their respective row and column.

drug-target interactions from a list of US Food and Drug Administration (FDA)-approved and experimental drug-target links. In vitro assays confirmed biological activity in five of the predicted drugs, specifically montelukast, diclofenac, simvastatin, ketoconazole, and itraconazole, on estrogen receptors or dipeptidyl peptidase-4 with IC_{50} values between 0.2 and 10 μM. They showed that simvastatin and ketoconazole had anticancer properties when introduced to the MDA-MB-231 breast cancer cell line in MTT assays. The confirmation of these in silico screens by in vitro methods is encouraging because it demonstrates that ML methods can lead to biologically relevant conclusions.

Yamanishi and colleagues (11) developed a drug-target interaction prediction engine for four different protein classes, i.e., enzymes, ion channels, G protein–coupled receptors, and nuclear receptors, using drug structural similarity, target protein sequence similarity, and drug-target interaction topology. Chemical structure information was collected from the KEGG LIGAND (12) database and analyzed using SIMCOMP (13), a computational tool for calculating global similarity scores based on the size of common substructures. Target protein sequences were collected from KEGG GENES (12), and similarity was calculated using Smith-Waterman scores (7). Drug-target information for the four protein categories was collected from KEGG BRITE (12), BRENDA (14), SuperTarget (15), and DrugBank (16). The authors first conducted a correlation analysis between their inputs, drug structure similarity, and protein target sequence similarity along with their outputs and drug-target interaction network topology, finding a strong relationship and indicating that the data can be useful for prediction (11). They then used the drug structure similarity and protein target sequence similarity features to predict unknown drug-target interaction networks using three separate statistical methods—a nearest profiles method, a weighted profile method, and a bipartite graph learning method that they had developed. Analysis of the network models demonstrated the ion channels to be greatly interconnected, indicating low specificity with respect to ligands, while the other three protein categories had significantly fewer connections and more individual clusters, demonstrating their greater specificity with respect to ligands. The AUC values presented for the Yamanishi et al. models (11) are misleadingly good, as the highest sensitivity for any of the modeling methods between all protein classes is 0.574 (i.e., not very good), while the corresponding AUC value is 0.904 (i.e., excellent). A precision recall curve (PRC), which graphs the precision of a model against its recall (17), would be more representative of the predictive power of the modeling method because it accounts for the unequal distribution of instances between the two classes being compared. [The precision of a model is the number of true positives it predicts over the sum of the true positives and the false positives (18). Recall is the number of true positives over the sum of the true positives and false negatives (18).] Nevertheless, the bipartite graphing method developed by Yamanishi et al. (11) has proven to be better at predicting novel drug-protein interactions in the data set, as shown by the gain in specificity when compared to the other methods. The success of the bipartite graphing method over other modeling methods in both works (6, 11) suggests that this architecture is particularly well suited for the prediction of drug-target interactions. The best results achieved by Yamanishi et al. (11) were for enzyme proteins (sensitivity of 0.574) and the worst results for nuclear receptor proteins (sensitivity of 0.148). Xiao and colleagues (8) built a model that is designed to handle the class imbalance presented in the previous work and in many similar studies. They employed a neighborhood cleaning rule and synthetic minority oversampling to balance the number of positive and negative samples in the data set. Neighborhood cleaning rules and synthetic minority oversampling are methods used to balance data sets whose distribution of instances in each class is not equal (19). Xiao et al. (8) then applied SVM to build a drug-target prediction tool for each protein class from the balanced data set. The least accurate model, predicting drug–ion channel interactions, was still very good (accuracy of 88.78%). The model predicting

interactions between drugs and nuclear receptors had the highest accuracy (93.02%), which is interesting given that the model developed by Yamanishi and colleagues (11) performed most poorly on this category of proteins. Both models performed relatively poorly in the prediction of interactions between drugs and ion channels—this was the worst category in the Xiao et al. (8) study and arguably the worst in Yamanishi et al.'s (11)—which is an interesting observation given the comment made by Yamanishi and coauthors (11) about the lack of ligand binding specificity in ion channel proteins.

COMBINATION THERAPIES

CTs are effective in treating complex diseases. They are helpful in fighting resistant disease strains (20–22) and tumors (23–25) because these therapies can attack diseases at multiple fronts, leading to synergistic interactions. Nevertheless, CT can be associated with increased toxicity. The combinatorial explosion resulting from the incredibly large and growing number of drugs necessitates the development of in silico methods for selecting optimized CTs with minimal toxicity.

AI-based models have been developed for predicting synergistic drug combinations in many ailments, including infectious diseases (20–22) and cancers (23, 24). General, non-disease-specific CT predictors have also been developed (25–28). The models vary greatly in their inputs, outputs, and architectures. Disease-specific models employ data specific to the patient's case such as genomics or proteomics, while general combination predictors make recommendations using only information related to drugs and their targets.

Combination Therapies for Infectious Diseases

Mason and colleagues (20) developed a model for predicting synergism of antimalarial drugs from data on experimental combination screens of drugs and molecular structures utilizing a RF architecture with high accuracy. RF is a tree-based algorithm commonly used for classification and regression models (29). The model was applied to a set of 253 drug combinations, of which it predicted 20 would be synergistic (20). The 20 drug combinations were then tested experimentally, and 9 of the combinations demonstrated synergism (20). The highest-ranked combinations included apicidin with virginiamycin S1, with a 62% expected probability of synergism. It is encouraging to see an in silico model confirmed by in vitro methods.

AI-assisted drug combination selection strategies have been developed for the treatment of human immunodeficiency virus (HIV) (21, 22). HIV treatment is complicated by the constant introduction of drug-resistant mutations of viral proteases and reverse transcriptase in the virus. This phenomenon forces treatment regimens to be periodically altered to address changes in the viral genetics and prevent the viral load from growing. The toxic load contributed by the drug must also be minimized. Computational approaches relying on ML were developed to keep up with changes in these proteins (21, 22). Wang et al. (21) developed a model predicting the virologic response to combinations of HIV medications. Three ML methods, artificial neural networks (ANNs), RF, and SVM, were used to predict the virologic response to a combination of drugs taking into consideration the viral mutation data and some added clinical biomarkers. [ANNs are mathematical models that are inspired by the brain and used for ML; they consist of layers of neurons and synapses that connect the neurons and apply weights to their outputs in the process of moving values through the network, eventually leading to outputs (30) (**Figure 4**).] Accuracy of results is assessed by the R^2 value between actual and expected results. The RF-based model developed by Wang and colleagues (21) achieved the highest individual accuracy with an R^2 value of 0.707. Even better results (R^2 of 0.728) were achieved by applying a committee approach,

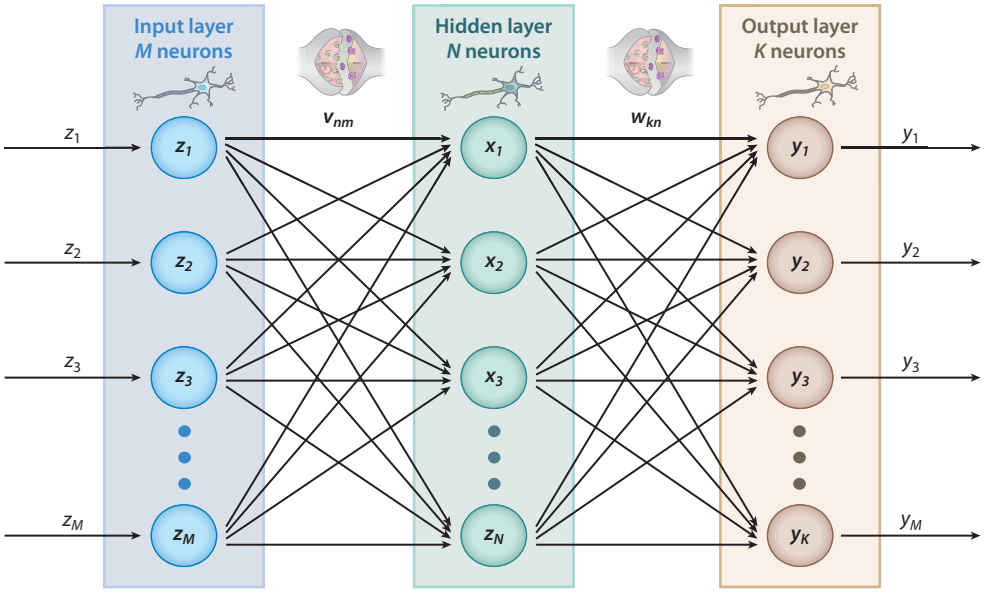


Figure 4

Diagram of artificial neuron (perceptron) processing. The circles labeled z_i, x_i, and y_i represent the neurons, or calculation centers, and the arrows represent synapses, which pass information between neurons and apply weights to the outputs of neurons. Mathematical functions, like tangent, are applied to the input values of the neurons. Synapses multiply the outputs of the neurons by weights. The weights applied by these synapses are constantly adjusted to minimize error until a minimum degree of error is achieved. The middle layer, or the hidden layer, is the source of the black box effect; this effect makes it difficult to interpret the reasoning of models because the inputs have already passed through a layer that performed mathematical operations on initial values, and the outputs are the result of another set of computations performed by the hidden layer on the already altered values. Deep learning often relies on ever-larger numbers of hidden layers, which leads to greater abstraction and decreased transparency. Figure adapted from Reference 32.

considering only treatment suggestions agreed upon by multiple models, to the outputs of all three models, demonstrating the gain in predictive power offered by a concordance of outputs (21). Prosperi and colleagues (22) developed a similar model, comparing three AI techniques to predict the response to antiretroviral treatment of 3,143 HIV patients in the EuResist database in terms of their viral load, immune cell count, and viral mutation profile (31). They compared the performance of a rule-based model to two ML models, one utilizing logistic regression and another RF. The best results came from the ML approaches, with RF being slightly better (AUC of 0.77), although logistic regression was very similar (AUC of 0.76). The rule-based system alone had notably worse results (i.e., a maximum AUC of 0.70). When the outputs of the rule-based system were used as inputs for the logistic regression model, the resulting AUC was comparable to the pure ML methods. We question whether the reported increase in accuracy has anything to do with the rule-based portion of the model, as it achieves similar results to logistic regression without the rule-based system. The higher performance of the ML methods over the rule-based system, with respect to prediction accuracy, supports the application of such nontransparent methods in reducing the dimensionality of complex biological data for application in drug treatment selection as an aid or form of second opinion for doctors to utilize.

Combination Therapies for Cancers

CT prediction engines for oncological decision support were developed using ML methods (23, 24). They focus on leveraging biological data relating to the genetics and/or proteomics of the disease integrated with chemical information, such as QSAR descriptors, to predict whether a drug combination will be synergistic. Li and coauthors (24) developed an ML model utilizing RF architecture in a binary classification, predicting the synergism of two drug combinations on specific cell lines. They used 18 features describing the drugs' physiochemical properties, target network properties, and pharmacogenomic properties. The first two feature categories, physiochemical and target network properties, accounted for three total features (CT features). The final category, accounting for 15 features, was represented using gene-expression profiles after drug perturbation on the DLBCL cell lines. Li et al. (24) evaluated performance as a function of input data and found that integrating all forms of available data led to the best results; nevertheless, the 15 differential gene expression features led to higher accuracy on their own than did the physicochemical and target network features on their own. Integration of all features led to a receiver operating characteristic (ROC) of 0.89, while the differentially expressed genes alone achieved an ROC of 0.83, and CT features alone achieved an ROC of 0.73. They applied the model developed from integrated data to an independent data set of drug combinations, predicting 28 synergistic combinations, only three of which (azacitidine with thalidomide, carmustine with streptozocin, and imatinib with paclitaxel) had been previously reported in the literature. The larger predictive contribution of the differentially expressed genes suggests that pharmacokinetic properties have more to do with the prediction of synergism than physicochemical or target network properties, but the fact that the model performed best with all available data supports the application of a global systems biology approach in which the model benefits from having a more complete picture.

Xia and colleagues (23) developed a different method for the prediction of synergism, and antagonism, between oncological drugs. They used a feed-forward neural network in the prediction of a scalar metric describing the degree of growth inhibition between the two drugs on the cell line. This differs from the previous synergy models because it is a regression problem, predicting a continuous value instead of a category. Positive values describe synergism, while negatives describe antagonism. Two categories of data were applied: chemical structure data, in the form of QSAR descriptors and molecular fingerprints for 104 drugs, and cell-line descriptors, in the form of gene expression micro arrays, microRNA expression profiles, and protein abundance measurements for the NCI-60 cell lines. They used mean squared error (MSE), mean average error, and R^2 values to represent the model's accuracy. Best results were again achieved through the integration of all data forms. The authors also analyzed the model's accuracy as a function of inputs and found drug chemical descriptors to be far more important than cell line descriptors (23). The R^2 value of model predictions jumps from 0.1272 to 0.9005 when using only gene expression and one hot encoded drug name or only gene expression and drug chemical descriptors, respectively. One hot encoding describes a method in which a list of categories is converted to a set of columns, with 1 denoting the presence of that instance in a given category and 0 denoting its absence (**Figure 5**). The large jump in R^2 value contributed by adding chemical features contrasts with the small gain in R^2 value, 0.8892 to 0.9208, contributed by the representation of cell line features with protein concentrations, expression levels, and RNA expression levels instead of only one hot encoded molecular features. These results suggest that the chemical feature information has greater importance to the predictive power of the model, as the authors argue, but could also imply that one hot encoding represents cell line features better than it does chemical features for this task. Examples of models built using only cell line or only chemical descriptors would help to elucidate which form of data is truly most important. Given the high R^2 value achieved by Xia

Genes	TP53	EGFR	BRAF	BRCA1
TP53	1	0	0	0
EGFR	0	1	0	0
BRAF	0	0	1	0
BRCA1	0	0	0	1

0 = Gene absent
1 = Gene present

Figure 5

An example of one hot encoding transformation, a method by which categories or labels can be represented in a way that is useful for machine learning. On the left is a list of genes, and on the right is a table in which 1 illustrates the presence of that gene in that row and 0 represents the absence of that gene in that row.

and colleagues (23), it would be interesting to see how the model's predictions would compare to novel drug combinations and cell lines in vitro or how the synergism and antagonism predictions correlate to responses in patients.

The main issue with the abovementioned prediction methods is that their reasoning is not transparent to the operator, making alteration of the treatment decision difficult. ML methods often sacrifice some amount of transparency for gains in predictive performance. Janizek and colleagues (33) were able to develop an ML model utilizing extreme gradient-boosted (GB) trees to build a transparent model with notable predictive power. [Gradient boosting is an ML technique used to optimize tree-based architectures with respect to their loss of function by adding tree elements whose predictive contribution beats that of random chance one at a time (34).] Janizek et al. (33) set up a regression model for the prediction of a Loewe drug synergy score [defined by Preuer et al. (35)] using a total of 8,846 features, including gene expression levels of cell lines, QSAR descriptors, and molecular fingerprints of drugs achieving a rank correlation score of 0.70 ± 0.02 and a MSE of 0.519 ± 0.08. The authors then used a feature ranking method for tree-based architectures, TreeSHAP (36), which gives weights to each feature representative of their importance in building the tree, to order the descriptors and demonstrated that the ML model suffers almost no loss in performance if only the top 1,000 features are used (MSE drops by 1.73%). They also examined the top 100 ranked features and found that only 17 of them are related to gene expression profiles; the rest were all drug-related features. This is consistent with past findings that drug features are of the greatest importance for synergy predictions. It is difficult to compare these results to those for other models because the MSE is calculated on a different scale than in other models and an R^2 value between predicted and actual values is not provided. The rank correlation value is, however, quite high, supporting the utility of this modeling method, and the transparency is a definite positive. It has been demonstrated that synergism predictions are significantly more dependent on drug features, chemical descriptors, similarity metrics, and interaction networks than on patient features and genetic profiles (23).

Universal Combination Therapies

Xu and colleagues (25) have built a general drug synergism predictor not focused on any specific disease, which utilizes only information related to drugs and their targets. The model employs 50 features related to the drug combination's biological, chemical, and pharmacological similarities and interactions in the prediction of effective drug combinations from the Drug Combination Database (37). Three ML methods, stochastic gradient boosting (SGB), SVM, and naïve Bayes (NB), were applied to predict drug synergy from the biological, chemical, and pharmacological data of the drugs. The tenfold cross-validation of their performances in predicting drug combinations demonstrated that SGB significantly outperformed the other two modeling methods in all recorded statistical metrics (AUC 0.9519, ROC 0.9775, F_1 0.8979, Recall 0.8693, and

Precision 0.9292) (25). [F_1 score is a statistical metric represented by the following formula: $F_1 = 2(\text{Sensitivity})(\text{Recall})/(\text{Sensitivity} + \text{Recall})$.] These are by far the best results presented for any of the drug combination predictors, strengthening the case that synergism prediction is much more dependent on drug-related data than on patient-related data.

Sun et al. (26) have also developed an interesting CT prediction model that is not focused on any specific disease. They integrated gene expression data of single drugs from the Drug Combination Database (38) and the human disease pathway for each drug from the Molecular Signatures Database (39), with three different frequency metrics representing the significantly changed pathways of the combination of drugs. Each frequency metric was applied in separate modeling attempts, and two modeling methods were attempted: SVM and NB. NB is a statistical method that multiplies the probabilities of various represented categories and gives the most probable class as a prediction. The authors explored how the accuracy changes as a function of kernel type in the SVM architecture, achieving the best results with a Gaussian kernel: 69.1% classification accuracy (26). The application of Hadoop to parallelize computing processes decreases run times by about two to three times for all processes except for NB modeling, which is five times slower. Sun et al. (26) state that this is caused by the small sample space (i.e., 152 two-drug combinations) and the benefits of parallelization through Hadoop would become apparent for NB as well if the sample space was much greater. This would allow for calculations on a larger set of individual drugs and for combinations beyond two drugs (26). Their assumption may be true, but they do not provide an explanation or a reference. Nevertheless, the SVM-based method achieves accuracy comparable to that of other modeling methods, and the frequency metric they developed describing significantly changed pathways significantly improved the accuracy of SVM when compared to other similar frequency scores.

Li and colleagues (24) have employed several similarity scores to infer synergistic and antagonistic drug combinations. They developed a probabilistic model utilizing an altered Bayesian network, termed the probability ensemble approach, which infers drug combination efficacy from similarity scores calculated for each combination's chemical structure, ATC codes, SIDER database (40) terms, drug-target protein sequence, protein–protein interaction network, and chemical fingerprints with an AUC of 0.90 and predicted adverse effects with an AUC of 0.95. This is incredible accuracy, especially considering that there are only six feature vectors for each combination. Each feature vector is the result of several calculations, so they may have a large amount of embedded information. One weakness specific to this model is its dependence on SIDER search term similarity, a categorical variable that can sometimes be assigned subjectively. Zhao et al. (28) developed another model for the recommendation of drug combinations from similar information. It differs in that the modeling method aims only to maximize the parameters with respect to the F_1 score and the way that protein targets of the drug combinations are represented. They (28) multiplied vectors representing each drug's protein targets to get all possible protein combinations, as seen in **Figure 6**. None of these models use any patient data, which is both a weakness

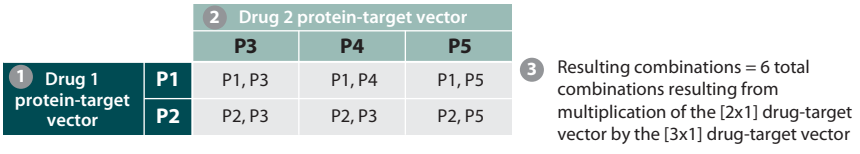


Figure 6

Protein-target vectors of drug combinations. The proteins associated with each individual drug are crossed to make multiple vectors, illustrating all possible combinations of target-protein pairs covered by a particular set of drugs. Figure adapted from Reference 28.

and a strength. These models, much like the treatments they are recommending, are most useful when used in combination with disease-specific models. They should be used as filters to choose between drug combinations that have already been matched and approved to the specifics of someone's disease or as screens for exploring new possibly synergistic drug combinations.

Combination Therapies for Drug–Drug Interactions

A discussion about decision support systems for selecting CTs would be incomplete without addressing DDIs. Predicting DDIs in silico is important to the drug treatment–selection and administration process because it can help to minimize adverse reactions and maximize dosage efficiency. This is especially true for the growing elderly population for whom the risk associated with polypharmacy is of growing concern (41) and for patients with complex diseases in which taking multiple drugs may be necessary. The probability of a patient taking multiple drugs at the same time grows with age, as the number of concurrent diseases a patient has increases (42). Complex diseases often require multifaceted drug regimens to target all problematic regions. In silico methods for the prediction of DDIs must be developed for use in clinics to prevent polypharmacy from contributing negatively to patient health.

Although DDIs are screened as part of the FDA approval process, many of them go unnoticed until after clinical trials due to the enormous number of possible combinations (43). There is currently no standard method to predict novel DDIs, but there is most definitely a necessity for the development and adoption of such methods. Several AI-based models were developed for the a priori detection of DDIs from biological, chemical, and pharmacokinetic data with high accuracy in the academic setting, but none have reached the clinical implementation stage. The prediction of DDIs using ML is typically accomplished through calculations related to the similarity in targets, pathways, and structures of two molecules. Few studies have tried to predict combinations of greater than two drugs, most likely due to the enormous amount of uncertainty contributed by the addition of each new entity and the exponential growth in computing power required with each added drug. Cheng & Zhao (43) have developed ML-based models for the prediction of DDIs by integrating data related to the phenotypic, therapeutic, genomic, and structural similarity of drug-drug combinations using NB, SVM, and logistic regression ML techniques. SVM marginally outperformed the others with an AUC of 0.666. The other ML methods achieved comparable performance and had the advantage of being simpler, but the trend of higher accuracy at the cost of transparency to the operator persists. Utilizing five common ML techniques (classification trees, k-nearest neighbor, SVM, RF, and GB machine), Kastrin and colleagues (27) developed a number of models for predicting DDIs by calculating the probability of there being a link (i.e., interaction) between two nodes (i.e., drugs) from topological indices describing network similarities and semantic features describing therapeutic, chemical, structural, side effect, and textual classification similarities between drugs. They noted that the usual metric for describing the predictive power of a model, ROC, is incomplete and even misleading for binary classifications with class distribution biases. They instead reported that the metrics, precision, recall, F_1 measure, AUC, and area under the PRC are better for characterizing these models. They also applied each of the five modeling methods to five DDI databases [DrugBank (16), KEGG (44), NDF-RT (45), SemMedDB (46), and Twosides (47)] separately and compared the results of each modeling method on each database. The highest predictive power was observed for RF-based models on DDI data from DrugBank and Twosides. The weakness of this modeling method is its dependence on semantic and topological data instead of purely quantitative variables like QSAR descriptors. Errors can be contributed by human operators mislabeling or wrongly categorizing class variables.

DOSAGE CONTROL AND TIMING

Properly controlling dosage quantity and administration timing is imperative to personalizing drug regimens. An ideal dosage regimen keeps the concentration of drug in the body at a constant equilibrium above the minimum effective concentration but below the minimum toxic concentration (48). This ideal administration of the drug is referred to as zeroth order release (48). Computational models have been exploited to approach the ideal zeroth order release to maximize drug effectiveness while minimizing side effects.

Implementation in the clinic would require a combination of real-time sensors with software for the measurement and optimization of the pharmacokinetic properties at play. Houy & Le Grand (49) developed a model for the optimization parameters of temozolomide dosage and timing with respect to the reduction of tumor size using a Monte Carlo tree–inspired method. They used the pharmacokinetic and pharmacodynamic data of temozolomide to represent the drug and used the nadir of the absolute neutrophil count, the myelosuppressor effect of temozolomide, and physiological factors described by Panetta and coauthors (50) to describe toxicity. The protocol developed using these techniques and data resulted in average predicted tumor shrinkage of 7.66 times with a 95% confidence interval of 7.36–7.97. These models were not tested on real patients' data, but an approach must be developed to foster in vivo testing of such results.

Some models aim at predicting how drugs will interact with the various transporters and membranes involved in the absorption, distribution, metabolism, excretion, and toxicity of drugs (termed ADMET properties) to elucidate how a patient may respond to a drug. ML approaches to predict ADMET properties can be drastically improved by the development of a systematic method for feature selection and parameter optimization (51). Yang et al. (51) studied the benefits of using a genetic algorithm for feature selection and the conjugate gradient method for parameter optimization in building four models relevant to drug pharmacokinetics with an SVM architecture. The conjugate gradient method is a mathematical technique that iteratively computes the direction of greatest decrease in error until a minimum is found (52). A genetic algorithm finds an optimal set of features by scoring each feature vector, or row of features, based on its informational contribution to the problem being addressed and crossing these high-scoring features with others (53). It repeats this process until the accuracy of the model is no longer increasing, thus finding the set of parameter weights that leads to the smallest error attainable. Genetic algorithm feature selection and parameter optimization methods have been demonstrated to be especially useful for constructing SVM-based ML models (54, 55).

Yang et al. (51) designed four models for the prediction of separate systems with which drugs may interact and possibly effect on their path to a target. The four ML models developed by Yang and colleagues (51) include the (*a*) identification of P-glycoprotein (P-gp) substrates and nonsubstrates, (*b*) prediction of human intestinal absorption (HIA), (*c*) identification of compounds inducing torsades de pointes (TdP) (polymorphic ventricular tachycardia), and (*d*) prediction of blood–brain barrier (BBB) penetration. Data for the first three models, representing the chemical agents, class labels, and molecular descriptors, were taken from the study by Xue et al. (56). Data for the final BBB model were acquired by compiling molecules from studies conducted by Zhao et al. (57) and Li et al. (58) and calculating their chemical descriptors using an online calculator called PCLINET. The feature and parameter optimization techniques applied here led to prediction accuracy increases of 16.8% in P-gp, 10.2% in HIA, and 4.1% in TdP with respect to models also built using SVM but without the genetic algorithm for feature selection and the conjugate gradient method for parameter optimization (51). The BBB model was treated separately and more thoroughly. Yang et al. (51) compared the results of this model to three previously built models from the literature, showing significant improvement. The observed gain in accuracy

ranged from 13.1% to 20.8% when compared to methods from other works and was 11.3% when compared to SVM with the genetic algorithm for feature selection but without the conjugate gradient for parameter optimization (51). The authors demonstrated the power of applying an SVM architecture to the prediction of pharmacokinetic properties, with accuracies between 85.1% for P-gp and 96.8% for BBB, and the enormous gain that can come from applying these optimization techniques in an SVM architecture (51). The best accuracy in each case was achieved by applying the techniques illustrated here, but some of this may be because the classes they were predicting are heavily biased. None of the data sets had an even split of instances in the two categories they were choosing between, and in every case the accuracy was much higher for the class with more representatives. They could have used a metric like PRC to demonstrate the predictive power of the model independent of this class bias or used a technique like synthetic minority oversampling, demonstrated by Xiao and colleagues (8), to remedy this issue.

Lim et al. (59) developed a recurrent neural network model for the prediction of dosage response over time for general diseases, termed the recurrent marginal structural network (RMSM). The mathematical model depicts tumor volume, V(t), to be the product of its size at a previous point in time, V(t-1), and a coefficient that is the sum of four terms: (*a*) tumor growth rate; (*b*) radiation dosage, sensitivity, and time of exposure; (*c*) chemotherapy dosage and sensitivity; and (*d*) a bias term. Lim et al. (59) aimed to demonstrate the modeling mechanism's ability to estimate unbiased treatment response and multistep prediction performance, hence the variables representing time and dosage. They compared this model's ability to predict tumor volume against other known methodologies, finding theirs to have the lowest root mean squared error (RMSE) when compared to several benchmarks and state-of-the-art methods. The R-MSM model's RMSE value grew only slightly, from 0.92% to 1.02%, with time-related confounding growing from 0 to 10, demonstrating the model's ability to adjust to treatment changes as time progresses with negligible growth in error. The predictive performance of this model, as measured by RMSE, exceeded that of traditional methods by 80.9% and that of the state-of-the-art Bayesian method by 66.1%. The model is composed of two parts: It first encodes a representation of a given patient's current disease state and then uses a decoder to predict the treatment response of a selected drug regimen administered over a set length of time. The accurate prediction of tumor volume, or any other disease's response to drugs, with time is incredibly important for maximizing the effectiveness of drugs while reducing the degree of unnecessary toxicity by elucidating the degree of disease progression as a function of drug susceptibility, dosage, and the time of administration.

DISCUSSION

The application of AI as an aid in the selection and administration of personalized drug regimens is vital, especially when combating complex diseases in which many patients need completely novel treatment protocols. Modern medicine achieved its massive success using technologies such as vaccinations and antibiotics, which generally work without much adjustment from patient to patient. Precision and personalized medicinal approaches are meant to combat the diseases left behind—those for which there are no one-size-fits-all solutions. Only through the application of computational tools dependent on AI engines can we truly leverage the vast amount of biological, chemical, and medical data that need to be integrated to fully understand and fight complex diseases such as cancers. There is no AI that will function as a skeleton key, recommending treatments for all diseases. Instead, tools are being developed to handle specific tasks in the drug selection and administration process of each disease and each patient. Much like other mathematical approximation methods, the functions that AI modeling methods determine are most accurate when they are not extrapolated over large output areas. Limiting the prediction space of models helps to boost their

accuracy by decreasing the number of possible outputs or the range of outputs for classification and regression problems, respectively. Even greater accuracy is achieved through the democratization of modeling methods, e.g., committee approaches like those in Chang et al. (4) and Wang et al. (21), where outputs depend on the decisions or predictions of multiple accurate representations.

The complexity of the data and modeling methods relied upon in AI technologies often results in the development of a decision-making process completely uninterpretable to human operators. This is inevitable, as there is no conceivable way for us humans to consider and integrate every piece of appropriate biological, chemical, and medical information relevant to each patient in the design of a treatment protocol specific to their needs. If we could make sense of it, we would not need the AI.

The purpose of AI in drug treatment is to reduce the data to a size the doctor can interpret, giving clinicians access to information they could not previously consider. Transparent reduction processes would have to be simple enough for the operator to make sense of them. Unfortunately, this sort of simplification often comes at the cost of a reduction in prediction power. AI-based black box decision support solutions contribute greater accuracy but with less transparency. There has been a lot of anxiety over the use of so-called black box models because their decision-making process is not always rationalizable to humans; but this stressor should be alleviated if they prove to be successful and patient survival increases. Extensive trials to justify any medical black box model's predictive abilities are a must because of the implications of such software and the lack of interpretability they often come with. The models must be demonstrated to irrefutably improve outcomes for patients before their application in clinics. The introduction and wider-scale use of DL for the selection and optimization of drug treatments will make it even more difficult for models to achieve transparency. We must learn to trust the decisions and advice provided by our machines, much like children must learn to trust the decisions of their parents.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. 2016. A survey of current trends in computational drug repositioning. *Brief Bioinform*. 17(1):2–12
2. Menden MP, Iorio F, Garnett M, Mcdermott U, Benes CH, et al. 2013. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLOS One* 8(4):e61318
3. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. 2011. Validation of QSAR models—strategies and importance. *Int. J. Drug. Des. Discov.* 2(3):511–19
4. Chang Y, Park H, Yang H, Lee S, Lee K, et al. 2018. Cancer Drug Response Profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci. Rep.* 8:8857
5. Murnane K. 2016. What is deep learning and how is it useful? *Forbes*, April 1. **https://www.forbes.com/sites/kevinmurnane/2016/04/01/what-is-deep-learning-and-how-is-it-useful/#56be0bbbd547**
6. Cheng F, Liu C, Jiang J, Lu W, Li W, et al. 2012. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLOS Comput. Biol.* 8(5):e1002503
7. Smith TF, Waterman M. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* 147(1):195–97
8. Xiao X, Min JL, Lin WZ, Liu Z, Cheng X, Chou KC. 2015. iDrug-Target: predicting the interactions between drug compounds and target proteins in cellular networking via benchmark dataset optimization approach. *J. Biomol. Struct. Dyn.* 33(10):2221–33
9. Campillos M, Kuhn M, Gavin A, Jensen LJ, Bork P. 2008. Drug target identification using side-effect similarity. *Science* 321(5886):263–66

10. Dimitriadou A, Gogas P, Papadimitriou T, Plakandaras V. 2019. Oil market efficiency under machine learning perspective. *Forecasting* 1(1):157–68
11. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. 2008. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24(13):i232–40
12. Kanehisa M, Goto S, Hattori M, Aoki-Konishita KF, Itoh M, et al. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34(Suppl. 1):D354–57
13. Hattori M, Okuno Y, Goto S, Kanehisa M. 2003. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125(39):11853–65
14. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, et al. 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32(Suppl. 1):D431–33
15. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, et al. 2008. SuperTarget and Matador: resources for exploring drug–target relationships. *Nucleic Acids Res.* 36(Suppl. 1):D919–22
16. Wishart DS, Know C, Guo AC, Cheng D, Shrivastava S, et al. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36(Suppl. 1):D901–6
17. Brownlee J. 2018. How and when to use ROC curves and precision-recall curves for classification in Python. *Machine Learning Mastery*. **https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/**
18. Scikit-Learn. 2018. Precision-recall. *Scikit-Learn*. **https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html**
19. Al Abdouli NO, Aung Z, Woon WL, Svetinovic D. 2015. Tackling class imbalance problem in binary classification using augmented neighborhood cleaning algorithm. In *Information Science and Applications: Lecture Notes in Electrical Engineering*, ed. K Kim, pp. 827–34. Berlin/Heidelberg: Springer
20. Mason DJ, Eastman RT, Lewis RPI, Stott IP, Guha R, Bender A. 2018. Using machine learning to predict synergistic antimalarial compound combinations with novel structures. *Front. Pharmacol.* 9:1096
21. Wang D, Larder B, Revell A, Montaner J, Harrigan R, et al. 2009. A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy. *Artif. Intell. Med.* 47(1):63–74
22. Prosperi MC, Altmann A, Rosen-Zvi M, Aharoni E, Borgulya G, et al. 2009. Investigation of expert rule bases, logistic regression, and non-linear machine learning techniques for predicting response to antiretroviral treatment. *Antivir. Ther.* 14(3):433–42
23. Xia F, Shukla M, Brettin T, Garcia-Cardona C, Cohn J, et al. 2018. Predicting tumor cell line response to drug pairs with deep learning. *BMC Bioinform.* 19(Suppl. 18):486
24. Li X, Xu Y, Cui H, Huang T, Wang D, et al. 2017. Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif. Intell. Med.* 83:35–43
25. Xu Q, Xiong Y, Dai H, Kumari KM, Xu Q, et al. 2017. PDC-SGB: prediction of effective drug combinations using a stochastic gradient boosting algorithm. *J. Theor. Biol.* 417:1–7
26. Sun Y, Xiong Y, Xu Q, Wei D. 2014. A hadoop-based method to predict potential effective drug combination. *Biomed. Res. Int.* 2014:196858
27. Kastrin A, Ferk P, Leskos B. 2018. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PLOS One* 13(5):e0196865
28. Zhao XM, Iskar M, Zeller G, Kuhn M, van Noort V, Bork P. 2011. Prediction of drug combinations by integrating molecular and pharmacological data. *PLOS Comput. Biol.* 7(12):e1002323
29. Tsigelny IF. 2019. Artificial intelligence in drug combination therapy. *Brief Bioinform.* 20:1434–48
30. Aaron. 2015. Everything you need to know about artificial neural networks. *Medium*, Dec. 28. **https://medium.com/technology-invention-and-more/everything-you-need-to-know-about-artificial-neural-networks-57fac18245a1**
31. EuResist. 2019. *EuResist prediction engine*. Database, EuResist, Rome, accessed May 8, 2009. **http://engine.euresist.org**
32. Kose U. 2018. An ant-lion optimizer-trained artificial neural network system for chaotic electroencephalogram (EEG) prediction. *Appl. Sci.* 8(9):1613
33. Janizek JD, Celik S, Lee S-I. 2018. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. bioRxiv 331769. **https://doi.org/10.1101/331769**

34. Brownlee J. 2016. A gentle introduction to the gradient boosting algorithm for machine learning. *Machine Learning Mastery*, Sept. 9. **https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/**
35. Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. 2018. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 34(9):1538–46
36. Lundberg SM, Erion GG, Lee S-I. 2018. Consistent individualized feature attribution for tree ensembles. arXiv:1802.03888 [cs.LG]
37. Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X. 2014. DCDB 2.0: a major update of the drug combination database. *Database* 2014:bau124
38. Liu Y, Hu B, Fu C, Chen X. 2010. DCDB: drug combination database. *Bioinformatics* 26(4):587–88
39. Chen L, Li BQ, Zheng MY, Zhang J, Feng KY, Cai YD. 2013. Prediction of effective drug combinations by chemical interaction, protein interaction and target enrichment of KEGG pathways. *Biomed. Res. Int.* 2013:723780
40. Kuhn M, Letunic I, Jensen LJ, Bork P. 2016. The SIDER database of drugs and side effects. *Nucleic Acids Res.* 44(D1):D1075–79
41. Hajjar ER, Cafiero AC, Hanlon JT. 2007. Polypharmacy in elderly patients. *Am. J. Geriatr. Pharmacother.* 5(4):345–51
42. Charlesworth CJ, Smit E, Lee DSH, Alramadhan F, Odden MC. 2015. Polypharmacy among adults aged 65 years and older in the United States: 1988–2010. *J. Gerontol. A Biol. Sci. Med. Sci.* 70(8):989–95
43. Cheng F, Zhao Z. 2014. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J. Am. Med. Inform. Assoc.* 21(e2):e278–86
44. KEGG. 2019. *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Database, Kyoto, Japan. **https://www.genome.jp/kegg/kegg1.html**
45. US Natl. Lib. Med. 2019. *NDFRT (National Drug File—Reference Terminology)—Synopsis*. Drug File Database, US Natl. Lib. Med., US Natl. Inst. Health, Bethesda, MD. **https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/**
46. US Natl. Inst. Health. 2017. *Semantic Knowledge Representation*. SKR Database, US Natl. Inst. Health, Bethesda, MD. **https://skr3.nlm.nih.gov/**
47. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. 2012. Data-driven prediction of drug effects and interactions. *Sci. Transl. Med.* 4(125):125ra31
48. Siegel RA, Rathbone MJ. 2012. Overview of controlled release mechanisms. In *Fundamentals and Applications of Controlled Release Drug Delivery*, ed. J Siepmann, RA Siegal, MJ Rathbone, pp. 19–43. Boston: Springer
49. Houy N, Le Grand F. 2018. Optimal dynamic regimens with artificial intelligence: the case of temozolomide. *PLOS One* 13(6):e0199076
50. Panetta JC, Kirstein MN, Gajjar AJ, Nair G, Fouladi M, Stewart CF. 2003. A mechanistic mathematical model of temozolomide myelosuppression in children with high-grade gliomas. *Math. Biosci.* 186(1):29–41
51. Yang SY, Huang Q, Li LL, Ma CY, Zhang H. 2009. An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs. *Artif. Intell. Med.* 46(2):155–63
52. Hestenes MR, Stiefel E. 1952. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* 49(6):409–36
53. Mallawaarachchi V. 2017. Introduction to genetic algorithms—including example code. *Towards Data Science*, Jul. 7. **https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3**
54. Huang C-L, Wang C-J. 2006. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* 31(2):231–40
55. Frohlich H, Chapelle O, Scholkopf B. 2003. Feature selection for support vector machines by means of genetic algorithms. In *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, *Sacramento, CA, November 5*, pp. 142–48. New York: IEEE

56. Xue Y, Li ZR, Yap CW, Sun LZ, Chen X, Chen YZ. 2004. Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. *J. Chem. Inf. Comput. Sci.* 44(5):1630–38
57. Zhao YH, Abraham MH, Ibrahim A, Fish PV, Cole S, et al. 2007. Predicting penetration across the blood–brain barrier from simple descriptors and fragmentation schemes. *J. Chem. Inf. Model.* 47(1):170–75
58. Li H, Yap CW, Ung CY, Xue Y, Cao ZW, Chen YZ. 2005. Effect of selection of molecular descriptors on the prediction of blood—brain barrier penetrating and nonpenetrating agents by statistical learning methods. *J. Chem. Inf. Model.* 45(5):1376–84
59. Lim B, Alaa A, van der Schaar M. 2018. *Forecasting treatment responses over time using recurrent marginal structural networks*. Poster presented at the 32nd Conference on Neural Information Processing Systems, Montreal, Canada
60. Natl. Cent. Biotechnol. Inf. 2019. *Acitretin*. PubChem Database, Natl. Cent. Biotechnol. Inf., US Natl. Inst. Health, Bethesda, MD. **https://pubchem.ncbi.nlm.nih.gov/compound/5284513**
61. Natl. Cent. Biotechnol. Inf. 2019. *Cetirizine*. PubChem Database, Natl. Cent. Biotechnol. Inf., US Natl. Inst. Health, Bethesda, MD. **https://pubchem.ncbi.nlm.nih.gov/compound/2678**