Send Orders for Reprints to reprints@benthamscience.net

Current Bioinformatics, xxxx, *xx,* x-x

REVIEW ARTICLE

Machine Learning and Deep Learning Strategies in Drug Repositioning

Fei Wang[1], Yulian Ding[1], Xiujuan Lei[2], Bo Liao[3] and Fang-Xiang Wu[1,4,*]

[1]*Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, Canada;* [2]*School of Computer Science, Shaanxi Normal University, Xi'an, China;* [3]*School of Mathematics and Statistics, Hainan Normal University, Haikou, China;* [4]*Department of Mechanical Engineering and Department of Computer Science, University of Saskatchewan, Saskatoon, Canada*

ARTICLE HISTORY

Received: April 29, 2021
Revised: August 15, 2021
Accepted: September 08, 2021

DOI:
10.2174/1574893616666211119093100

Abstract: Drug repositioning invovles exploring novel usages for existing drugs. It plays an important role in drug discovery, especially in the pre-clinical stages. Compared with the traditional drug discovery approaches, computational approaches can save time and reduce cost significantly. Since drug repositioning relies on existing drug-, disease-, and target-centric data, many machine learning (ML) approaches have been proposed to extract useful information from multiple data resources. Deep learning (DL) is a subset of ML and appears in drug repositioning much later than basic ML. Nevertheless, DL methods have shown great performance in predicting potential drugs in many studies. In this article, we review the commonly used basic ML and DL approaches in drug repositioning. Firstly, the related databases are introduced, while all of them are publicly available for researchers. Two types of preprocessing steps, calculating similarities and constructing networks based on those data, are discussed. Secondly, the basic ML and DL strategies are illustrated separately. Thirdly, we review the latest studies focused on the applications of basic ML and DL in identifying potential drugs through three paths: drug-disease associations, drug-drug interactions, and drug-target interactions. Finally, we discuss the limitations in current studies and suggest several directions of future work to address those limitations.

Keywords: Machine learning, deep learning, drug repositioning, drug-disease association, drug-drug interaction, drug-target interaction.

1. INTRODUCTION

In the traditional pharmaceutical industry, introducing a new drug into the market is very costly and time-consuming. Expenditure of about 1 billion US dollars and time period of ten years are common [1]. The related budgets are still increasing rapidly. In the traditional drug discovery pipeline, four major procedures are essential: drug discovery, preclinical experiments, clinical trials, and regulatory approval [2], as shown in Fig. (**1**). Several thousands of small compound candidates are typically studied to develop one new drug. However, in many projects, no drug can be taken to the market successfully.

Drug repositioning approaches have been proposed to identify novel treatments for existing drugs in order to save time, reduce cost, and improve the possibility of success. The safety and other properties of existing drugs have been studied clearly so that the pre-clinical periods can be reduced significantly. Some successful drugs have been identified to serve as novel treatments for different diseases, and have approved by the United States Food and Drug Administration (FDA), such as sildenafil, thalidomide, zidovudine, minoxidil, and celecoxib [3]. Those drugs are generated by two types of drug repositioning approaches, which are phenotypic screening and target-based approaches [4]. In the first decade of the 21st century, 45 small compounds were proposed by those two types of approaches, 28 of which were identified by phenotypic screening [5, 6].

However, the traditional drug repositioning approaches still have some limitations. In phenotypic screening, small animal models and cell-based models are necessary. The robustness and relevance of models influence the success of screening [7]. In target-based approaches, the experiments are based on assays, and the number of effective drug targets is limited [8]. Computational drug repositioning approaches have been proposed to address those limitations. Based on biological data, various algorithms and applications have been proposed to identify novel treatments for existing drugs.

Machine learning (ML) technologies have been applied in many computational fields and achieved good performance in solving regression, classification, and clustering problems. The concept of "machine learning" was proposed by Alan Turing in the 1950s [9]. They are useful tools to identify potential drugs in drug discovery. Deep learning and

*Address correspondence to this author at the Division of Biomedical Engineering, College of Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, Canada; Tel: 1(306) 966-5280; E-mail: fangxiang.wu@usask.ca

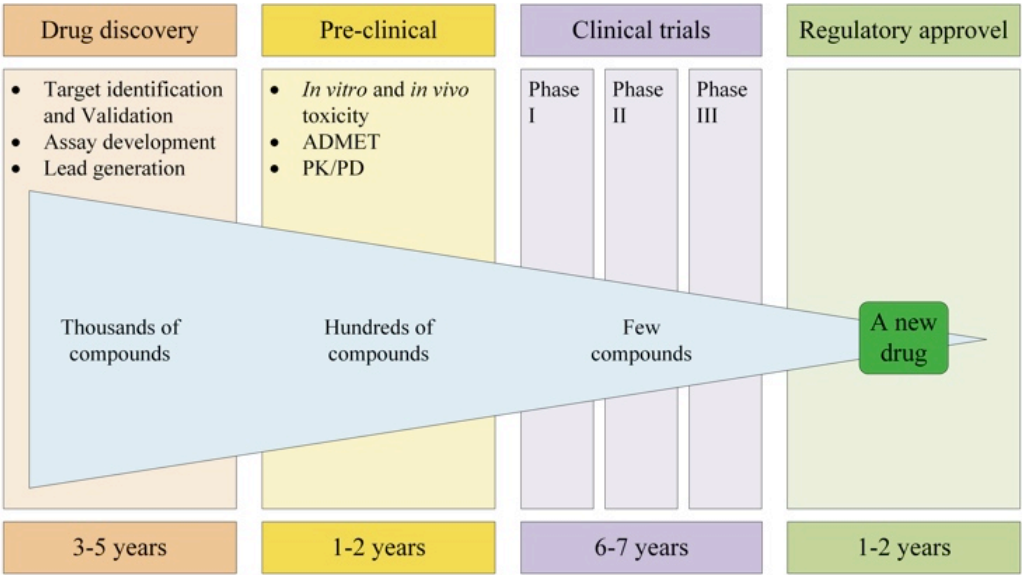


Fig. (1). The drug discovery pipeline.

basic ML are two classes of ML. The basic ML strategies, such as basic neural network (NN) [10-20], decision tree [21, 22], random forest (RF) [22-31], *k*-nearest neighbor (KNN) [22, 32], random walk (RW) [33-42], support vector machine (SVM) [15, 25, 43-47], and shallow autoencoder [19, 26, 41, 48-50], have shown their successful usages in predicting potential drug-disease associations (DDAs), drug-drug interactions (DDIs), and drug-target interactions (DTIs). Those associations and interactions help identify novel treatments for existing drugs. Many researchers apply ML methods to extract drug features, disease features, and target features from public databases and make predictions based on those feature vectors [10, 22, 32, 43, 44, 51]. Other researchers employ ML methods to predict potential missing links on the drug-disease heterogeneous network [33]. The networks are based on known links and similarities. The inferred probability values are given to the unknown links which are not shown on the network.

Deep Learning (DL) has also been applied to drug repositioning recently; Wen *et al*. utilized a DL method to predict potential DTIs [52], which is the first DL application in predicting DTIs. After that, many DL methods have been applied to predict potential DTIs, DDAs and DDIs, such as deep neural network (DNN) [15, 25, 42, 49, 53-60], convolutional network (CNN) [11, 14-17, 19, 27, 55, 58, 59, 61-65], recurrent neural network (RNN) [16, 55], and stacked autoencoder (SAE) [28, 30, 42, 66].

In the applications, many methods focus on predicting some novel DDAs, DDIs, and DTIs. DDAs provide essential information for drug repositioning [48]. Novel associations may reveal the treatments of existing diseases with new drugs.

DDIs reflect the relationships between two drugs. A drug may enhance the therapeutic efficacy of a drug and reduce the toxicity of another drug [67]. The predictions of DDIs help find some drug combinations that provide an effective treatment for a disease. Additionally, based on the "guilt-by-association" principle [68, 69], similar drugs may provide similar treatments.

Identifying DTIs is essential as it provides insights into the experimental design of drug discovery [70]. The targets are molecules that have proven associations with particular diseases [71]. Prediction of novel DTIs helps find novel usages of existing drugs.

According to the workflow of this review shown in Fig. (**2**), we first summarize some commonly used databases for drug repositioning purposes in Section 2. The most commonly used data types are drug features, disease features, target features, DDAs, DDIs, and DTIs [72]. The basic ML models and DL models are based on the feature vectors, associations, and interactions extracted from the databases. About 15 commonly used methods are introduced in Section 3. Then their latest applications in drug repositioning are systematically reviewed in Section 4. To make them clear, we divide them into three parts: the predictions of DDAs, DDIs, and DTIs. Finally, we discuss the limitations of those applications and some directions of future work in Section 5.

2. RELATED DATABASES AND PRE-PROCESSING STEPS

Drugs, diseases, and targets are key components for drug repositioning. Therefore, we first summarize some of the widely used databases for drug-, disease- and protein-centric information in Table **1**. Those data types include many feature types, such as drug chemical structures, disease phenotypes, and protein amino acid sequences.

Before employing ML and DL algorithms, two preprocessing steps are commonly adopted, including calculating similarities and constructing networks. The similarity is between two instances of the same type: drug-drug, disease-disease, and target-target. Various methods have been proposed to calculate the similarities, while some of the methods can be used in more than one type, as shown in Table **2**.

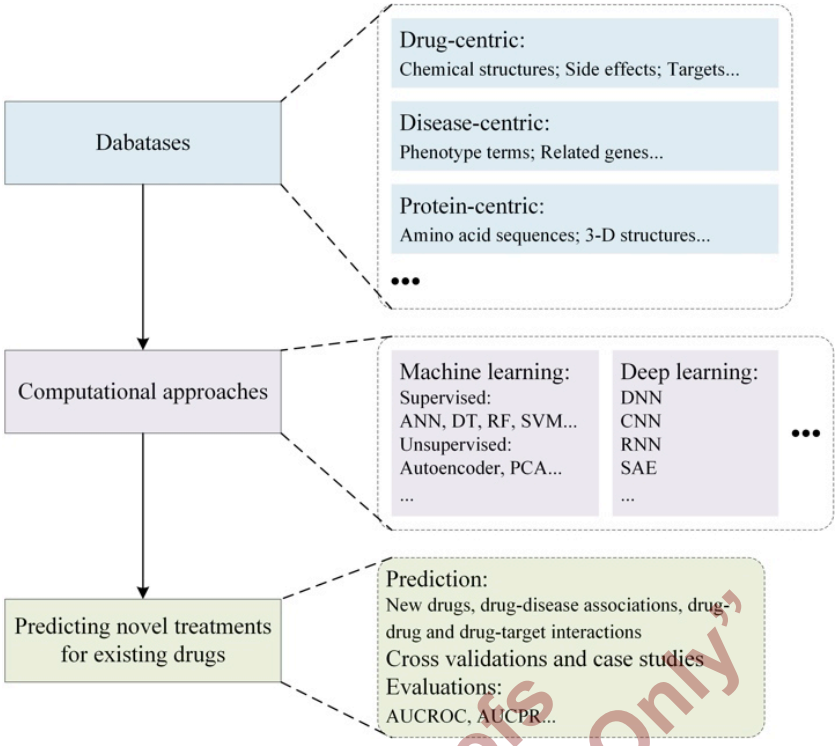


Fig. (2). The workflow of this review.

Table 1. Drug-, disease- and protein-centric databases.

Names	Descriptions	URLs
BRENDA	Drug target sequences and 3-D structures.	https://www.brenda-enzymes.org/
ChEMBL	Physicochemical properties of drugs.	https://www.ebi.ac.uk/chembl
CMap	Drug perturbation profiles.	https://clue.io/cmap
Comparative Toxicogenomics Database (CTD)	Drug-gene, gene-disease, drug-disease and gene-gene associations.	http://ctdbase.org/
DGIdb	Drug-related genes, Drug-gene annotations, interactions and potential drug ability database.	https://www.dgidb.org/
DisGeNet	Disease-related genes.	https://www.disgenet.org/
Drug target common (DTC) database	Drug-target interactions.	https://drugtargetcommons.fimm.fi/
ENCODE	Database of comprehensive parts list of functional elements in human genome.	https://www.encodeproject.org/
FAERS	Adverse event reports and medication error reports submitted to FDA.	https://www.fda.gov/drugs/ surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers
GEO	High throughput gene expression datasets.	https://www.ncbi.nlm.nih.gov/geo/
IUPHAR	Drug-target interactions.	https://www.guidetopharmacology.org/
KEGG	Databases dealing with genomes, biological pathways, diseases, drugs, and targets.	https://www.genome.jp/kegg/
LINCS	Dataset of transcriptional responses of human cells to chemical and genetic perturbation. 1.3 Million L1000 profiles and tools for their analysis.	https://lincsproject.org/
National Drug File Reference Terminology (NDF-RT)	Drug characteristics, including ingredients, chemical structure, dose form, physiological effect, mechanism of action, pharmacokinetics, and related diseases.	https://bioportal.bioontology.org/ontologies/NDFRT

(Table 1) Contd…

Names	Descriptions	URLs
NCI-DTP	Growth inhibition data.	https://dtp.cancer.gov/
OMIM	Human genes and genetic phenotypes.	https://www.omim.org/
Open Targets Platform	Comprehensive and robust data integration for access to and visualization of potential drug targets associated with the disease.	https://www.targetvalidation.org/
PubChem	More than 90 million compounds' chemical information along with their bio activities, gene and protein targets.	https://pubchem.ncbi.nlm.nih.gov/
SIDER	Adverse drug reactions, side effects and the indications of marketed medicines; information on marketed medicines and their recorded adverse drug reactions.	http://sideeffects.embl.de/
STRING	Protein-protein interactions, analysis, and networks.	https://string-db.org/
SuperTarget	Drug-target relations.	https://bioinformatics.charite.de/supertarget/
Therapeutic Target Database (TTD)	Dataset of known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information and the corresponding drugs directed at each of these targets.	http://db.idrblab.net/ttd/

Table 2. The feature types and similarities of drug-drug, disease-disease, and target-target associations.

Association Types	Feature Types	Similarity Methods/Tools
Drug-Drug	Chemical structure	CDKSim [73], SIMCOMP [74], Marginalized [75], Tanimoto [76], Spectrum and Lambda-k [77]
	ATC codes	ATCSim [78]
	Associated targets	Tanimoto [76], GIP [32]
	Side effects	Sider2 [79], Aers-bit and Aers-freq [80]
Disease-Disease	Phenotypes	SemFunSim [81], Separation [82]
	Ontologies	DoSim [83]
	Associated genes	GIB and PSB [84], ICod [85]
Target-Target	Amino acid sequences	Smith–Waterman algorithm [86], Spectrum and Mismatch [87]
	Ontologies	Semantic similarity [88]
	Associated drugs	GIP [32]

A heterogeneous network is a network with two or more types of instances. An example of a drug-disease heterogeneous network is shown in Fig. (**3**) [33]. It consists of a drug similarity network, a disease similarity network, and a drug-disease association network. The known drug-disease associations are used to connect the two similarity networks. In the similarity networks, the weights of interactions are based on the similarities. 5 different values of weights are used as examples in Fig. (**3**).

Some network-based methods are employed to identify the missing links in the heterogeneous network. Additionally, an adjacent matrix is determined from the network and it can be used to extract drug feature vectors and disease feature vectors.

3. BASIC MACHINE LEARNING AND DEEP LEARNING STRATEGIES

In this section, we illustrate the computing strategies of basic ML and DL, as shown in Table **3**. For basic ML, we discuss eleven commonly used methods. For DL, we introduce four types of deep neural networks (DNNs).

3.1. Basic Machine Learning Strategies

The basic idea of machine learning (ML) is to construct a model based on sample data. The models are used in a variety of applications, such as pattern recognition and drug repositioning. In this section, we introduce eleven widely used basic ML methods in drug repositioning, which are grouped into four categories: regression-based methods, ensemble methods, instance-based methods, and neural network methods.

3.1.1. Classification-based Methods

The classification-based methods are based on the linear combination of features to assign samples into two or more classes. The logistic regression and support vector machine are two typical classification-based methods, which are commonly used in binary classification problems of drug repositioning.

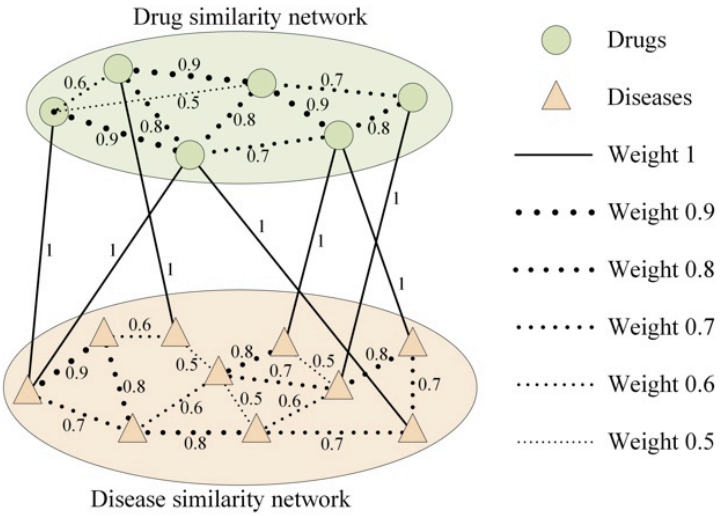


Fig. (3). An example of a drug-disease heterogeneous network. The solid lines denote the known drug-disease associations, and the weights of dotted lines denote the similarities. Six different weight values are exemplified. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Table 3. The introduced basic ML and DL strategies.

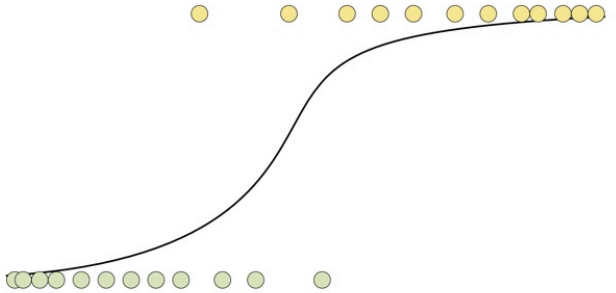
Types of Strategies	Commonly Used Strategies
Classification-based	Logistic regression Support vector machine
Ensemble Methods	Decision tree Bagging Boosting Random forest
Instance-based Methods	K-nearest neighbor K-means Random walk
Neural Network Methods	Basic neural network Basic autoencoder
Deep Learning Methods	Convolutional neural network Recurrent neural network Deep autoencoder Generative Adversarial network

Logistic Regression (LR) employs a logistic function to model a binary dependent variable. Most of the predictions of DDAs (or DDIs and DTIs) are binary classification problems. Therefore, the binary LR model has a dependent variable with two possible labels: "0" and "1". The log-odd for the value labeled "1" is a linear combination of independent variables. The probability of the variable labeled "1" varies between 0 and 1, that is a logistic function is used to convert log-odds to probability, as shown in Fig. (**4a**). A few researchers employ LR to predict potential drugs. Liu *et al*. utilized several ML models to predict novel DDAs, including LR [25].

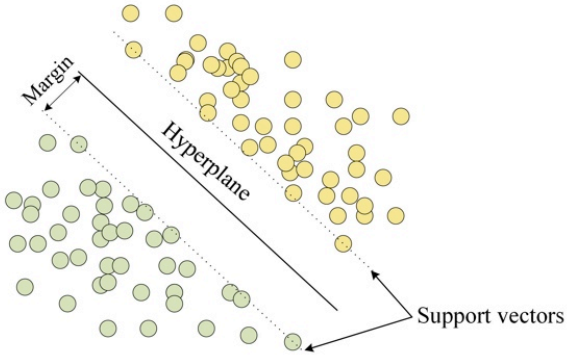
Support Vector Machine (SVM) is one of the most widely used classification algorithms [89]. When dealing with binary classification problems, SVM generates a hyperplane in the sample space. A good separation is achieved by the hyperplane with the largest distance to the nearest training sample of any class. The larger the distance is, the lower the error of the classifier is. An example of SVM for binary classification is shown in Fig. (**4b**). The SVM can be used to predict potential DDAs, DDIs, and DTIs [15, 22, 25, 43, 45, 46]. Beyond those, Zheng *et al*. employed the SVM algorithm to identify some satisfied reliable negative DDIs from unknown DDIs [44]. The known DDIs and reliable negative DDIs are utilized to predict potential DDIs.

3.1.2. Ensemble Methods

The ensemble methods combine multiple models to produce improved results of base models. In drug repositioning, many researchers use a decision tree as the base model and apply bagging and boosting methods to improve it. In the following, we mainly review decision tree, bagging, random forest, and boosting methods.



(a). Logistic regression



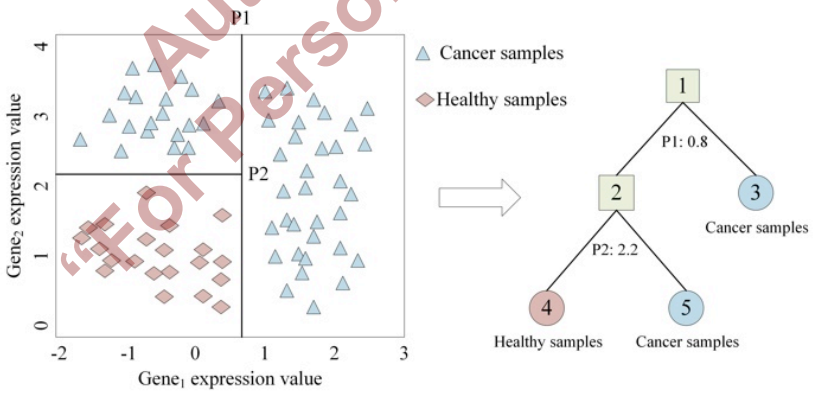
(b). Support vector machine

Fig. (4). Examples of logistic regression (**a**) and support vector machine (**b**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

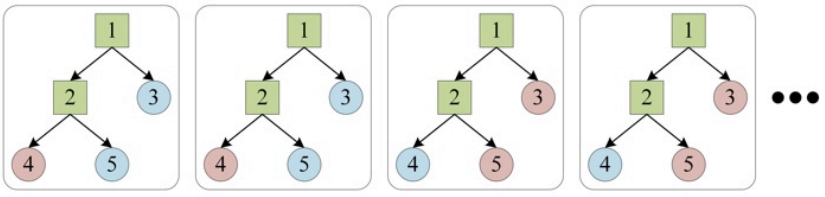
Decision Tree is used in many areas such as radar signal classification, medical diagnosis, and speech recognition [90, 91]. It is a tree structure model. Each internal node is a decision on an attribute, each branch is the outcome of a decision, and each leaf node is a class label Fig. (**5**). The paths from the root node to leaf nodes are classification rules. An example is shown in Fig. (**5a**), in which both cancer samples and healthy samples have two gene values. A decision tree model is constructed to distinguish cancer samples from healthy samples. In this review, the employment of decision tree model as a classifier is discussed for predicting potential drugs [21, 22].

Bagging is an abbreviation of "bootstrap aggregating." It is an ensemble algorithm to reduce variance and avoid overfitting [92]. It is often combined with other ML methods, such as decision trees. An example is shown in Fig. (**6a**). *n* datasets are generated from the original dataset by sampling along with replacement. Each dataset has the same sample size. A classifier is constructed in each subset. The voting of the outputs of all classifiers is the result of the bagging strategy. When processing regression problems, the result is the average of the outputs of all models.

Random Forest (RF) is an application of the bagging method in classification. It is a combination of decision trees in which each tree is constructed independently [93], as shown in Fig. (**5b**). It retains the benefits of decision trees while achieving better results by bagging samples [94]. It works well when dealing with biological datasets with a large number of features. Many researchers apply RF to predict potential drugs [21-30]. In those applications, the RF



(a). Decision Tree



(b). Random forest

Fig. (5). Examples of decision tree (**a**) and random forest (**b**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

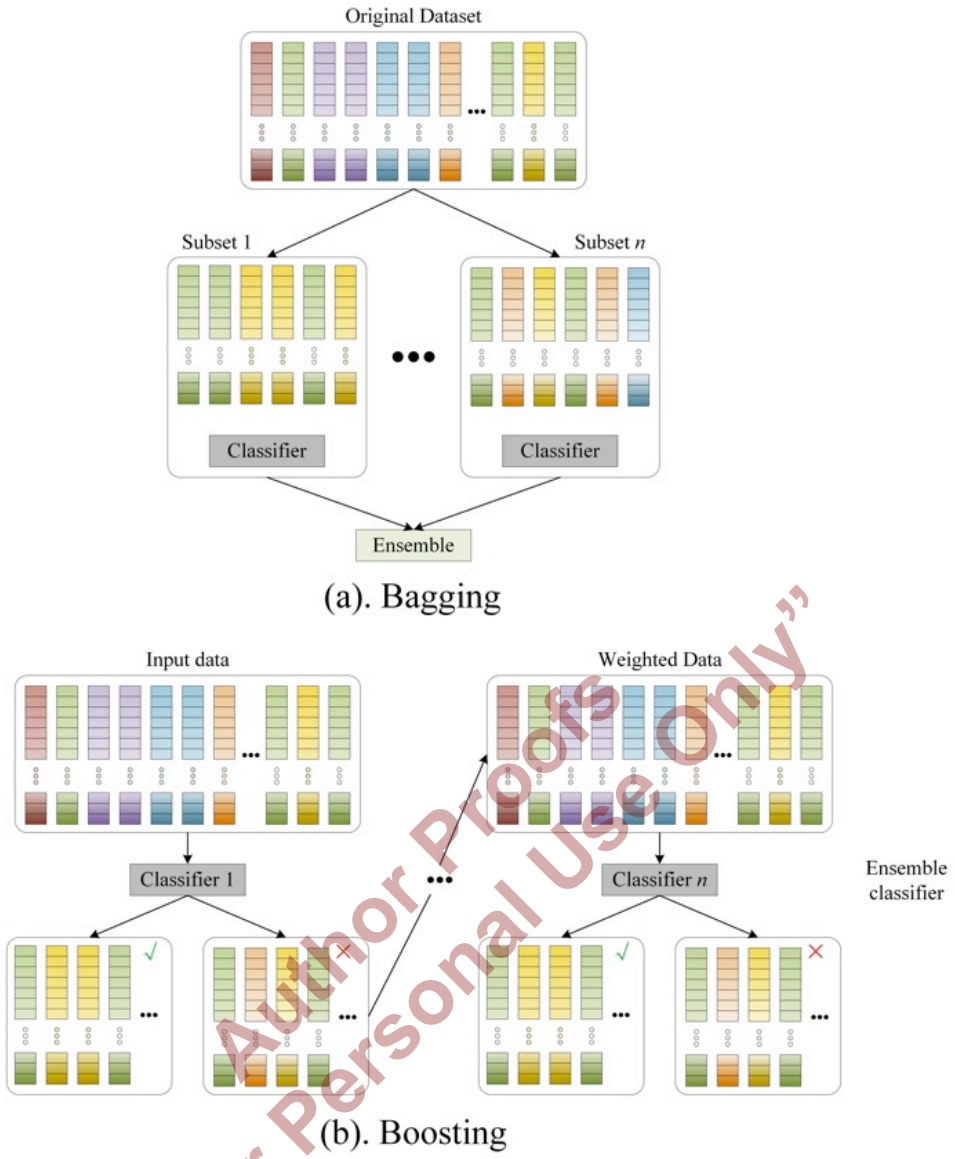


Fig. (6). The structures of bagging (**a**) and boosting (**b**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

model is a good classifier when processing vectors with thousands of features.

Boosting is another type of ensemble algorithm [95]. Most boosting algorithms consist of several classifiers in sequence. The first classifier classifies the training data. Then the misclassified data gain a higher weight, and correctly classified data lose weight. The second classifier works on the weighted data and updates the weights, as shown in Fig. (**6b**). The multiple weak classifiers can form a strong classifier *via* boosting.

The Adaptive Boosting (AdaBoost) [96] and Gradient Boosting [97] are two algorithms that use the boosting method. In AdaBoost, the outputs of the weak classifiers are combined into a weighted sum, while the weights are updated iteratively to adapt to the weak classifiers. In Gradient Boosting, the model is trained based on the residual between the true value and the predicted value of each sample. In predicting potential drugs, those algorithms are often combined with a decision tree or RF [29].

3.1.3. Instance-based Methods

The instance-based methods compare new instances with the training instances. This review mainly focuses on k-nearest neighbor, k-means clustering, and random walk.

K-nearest neighbor (KNN) is a typical instance-based method, either for classification or for regression problems [98]. Because KNN relies on distances to determine the nearest neighbors, a normalization process is useful to improve its accuracy, especially when the features vary in different scales. A commonly used distance metric is the Euclidean distance. An example of samples in 2-D space is shown in Fig. (**7a**). In a classification problem, a voting process is employed in the input sample's k nearest neighbors. The input sample is assigned to the class that has more votes

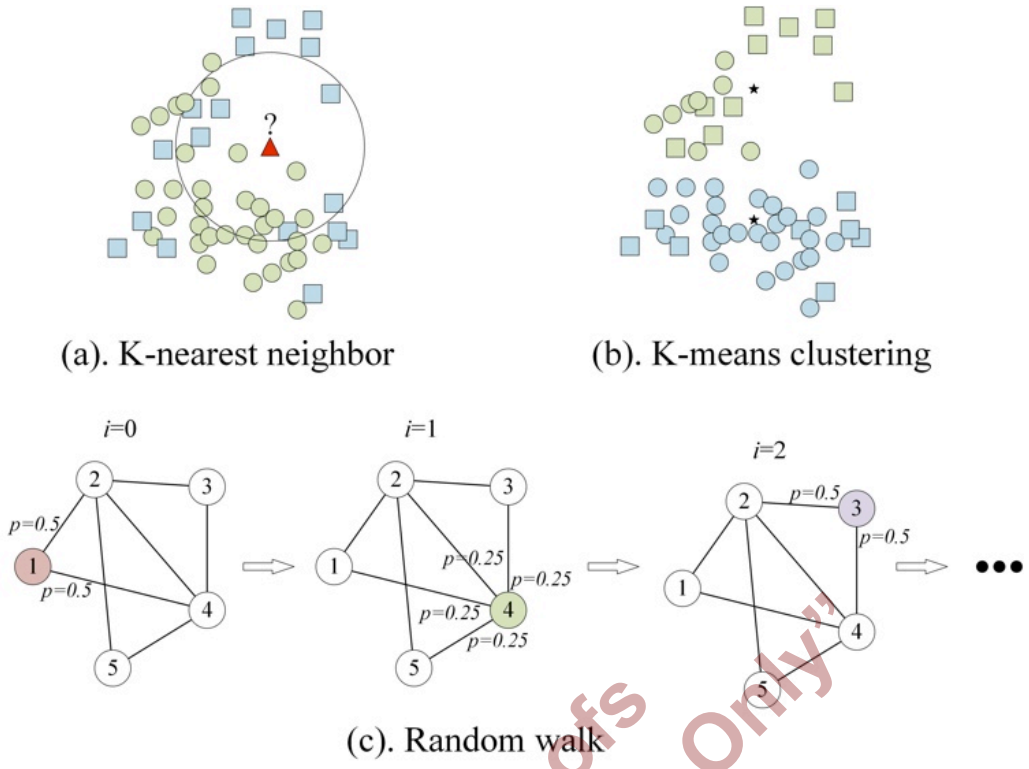


Fig. (7). Examples of k-nearest neighbor (**a**), k-means clustering (**b**), and random walk (**c**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

among the neighbors. When processing a regression problem, the input sample has an average value of its *k* nearest neighbors. Both types of problems are applicable in drug repositioning. In an earlier study [32], each known DDI had an intra-similarity, while the score of an unknown DDI was the average similarity of its *k* nearest known DDIs. In another study [22], KNN was applied to predict potential DDIs.

K-means Clustering (KMC) aims to cluster samples into *k* clusters. Each cluster has a center, and each sample belongs to the class whose center is the nearest center to the sample; then each center is updated according to the samples assigned to it, as shown in Fig. (**7b**). *k* is determined by users, and *k* samples are randomly identified as the initial centers of classes. After all the other samples are assigned to the nearest class, the centers are updated. Then the samples are assigned to the nearest classes iteratively. The algorithm is converged when assignments do not change significantly. In drug repositioning, KMC helps find the subsets of a dataset. Wang *et al*. utilized KMC to generate sub-types from cancer samples and identify a gene signature from each subset [99].

Random Walk (RW) is a stochastic process that the position of an instance in the (*i*+1)-th movement is only determined by its position in the *i*-th movement and a transition probability between those two movements, as shown in Fig. (**7c**). In similarity networks and heterogeneous networks, RW is a useful method to study the topological properties. In drug repositioning, many researchers used RW and its variations to predict potential drugs based on the drug-disease and drug-target heterogeneous networks [33-42].

3.1.4. Neural Network Methods

Neural networks are powerful models in machine learning. In the following, we mainly focus on basic neural networks and basic autoencoders, while deep networks are discussed in Section 3.2.

Basic Neural Network (NN) is a network method that contains three types of layers: input layer, hidden layer, and output layer [100]. The neurons in a layer are fully connected with those in the neighbor layers, as shown in Fig. (**8a**). Taking the neurons in the hidden layer for instance, the information is transformed as follows:

$$H^{Out} = \delta\left(W_H H^{In} + B_H\right) \quad (1)$$

Where, δ is the activation function in the hidden layer, H^{In} and H^{Out} are the inputs and outputs of the hidden layer, respectively. Meanwhile, the inputs of the hidden layer are the outputs of the input layer, and the outputs of the hidden layer are the inputs of the output layer. W_H and B_H are the weight matrix and bias vector of the hidden layer.

There are different activation functions, such as Sigmoid, TanH, eLU, ReLU, Leaky ReLU, and Softmax. The researchers can use any of them according to their requirements.

Many cost functions, which represent the differences between the predicted values and real values, are defined in applications. The cost function is used to optimize the values of parameters. One of the frequently used cost functions in

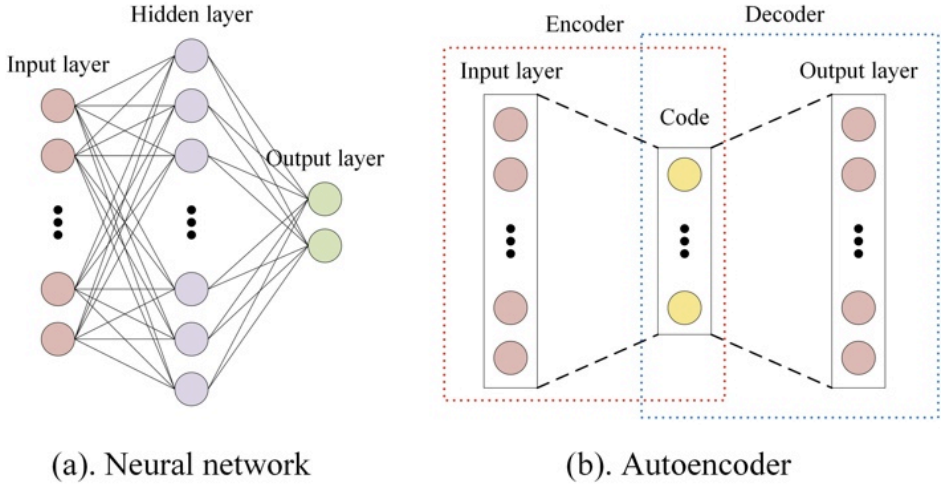


Fig. (8). The structures of basic neural network (**a**) and basic autoencoder (**b**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

processing binary classification problems is the binary cross-entropy cost function as follows:

$$Cost = -\frac{1}{n}\sum_{x}\left[y\ln(p)+(1-y)\ln(1-p)\right] \quad (2)$$

Where, n is the number of training samples, x is a training sample, and y is the label of x, while p is the prediction value. y has two possible values: "0" and "1".

In this review, the NN model is discussed in Section 4 for predicting the potential DDAs (DDIs or DTIs) [10-19]. The inputs of the NN are feature vectors extracted by different methods, and the outputs are the probabilities of the potential DDAs, DDIs, and DTIs.

Basic autoencoder is a type of NN that learns to copy its input to its output. The input layer and the output layer have the same number of neurons. The autoencoder has a code layer that describes a code to represent the input. It consists of two parts: an encoder maps an input to a code, and a decoder maps the code to an output. An example of a shallow autoencoder is shown in Fig. (**8b**). In drug repositioning, the autoencoder model is often utilized to reduce the dimensionality of feature vectors [19, 26, 41, 48]. Their dimensions are reduced from thousands to hundreds, and the predictions in the following processes become satisfying.

3.2. Deep Learning Strategies

The neural network with multiple hidden layers between the input layer and output layer is defined as a "deep neural network (DNN)," which underpins deep learning. The widely used convolutional neural network (CNN) [101], recurrent neural network (RNN) [102], Deep Autoencoder (DAE) [103], and generative adversarial network (GAN) [104] are different types of DNNs with different structures.

Convolutional Neural Network (CNN) utilizes several convolutional layers, pooling layers, and fully connected layers to form the model, as shown in Fig. (**9a**). The convolutional layer uses kernels to encode its input data [105]. In this layer, the widely used activation function is ReLU. The pooling layer aims to reduce the dimensionality of the data by integrating several neighbor neurons of one layer into a single neuron in the next layer. Max-pooling and average-pooling are two common types of pooling. Max-pooling transforms the maximum value among neighbor neurons of the prior layer to the next layer, while the average-pooling layer uses the average value instead. After several convolutional layers and pooling layers, a few fully connected layers are applied to generate the prediction results. CNN models can be employed to predict potential DDAs, DDIs and DTIs [11, 14-17, 19, 27, 55, 58, 59, 61].

Recurrent neural network (RNN) is a class of neural networks that the connections between neurons form a directed graph along a temporal sequence, as shown in Fig. (**9b**). The neurons at time t get inputs from other neurons in previous time steps. The calculation processes are as follows:

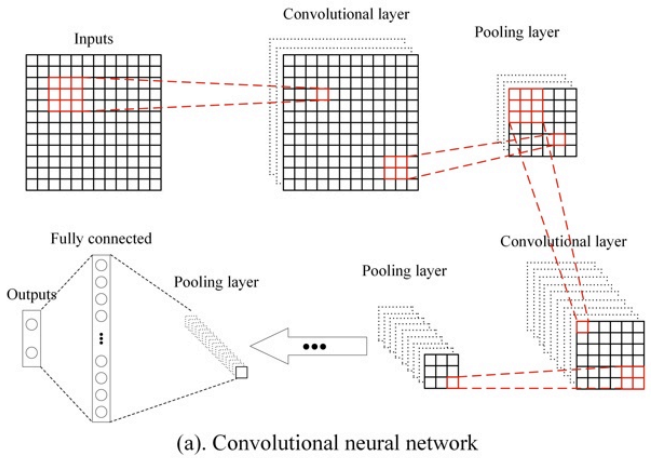
$$Y_t = g\left(VH_t + B_Y\right) \quad (3)$$

$$H_t = f\left(UX_t + WH_{t-1} + B_H\right) \quad (4)$$

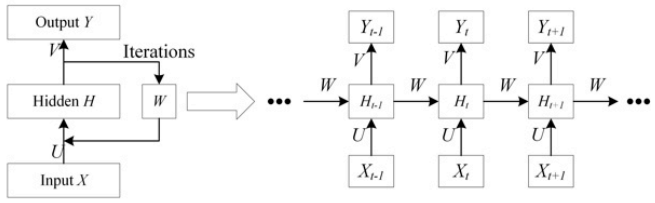
Where, the U, V, and W are weight matrices. B_Y and B_H are bias vectors. X_t, H_t, and Y_t are the matrices of the input layer, hidden layer, and output layer at time t, respectively. g and f are activation functions.

Deep autoencoder (DAE) is an autoencoder with multiple hidden layers, as shown in Fig. (**10a**). Both the encoder and the decoder consist of some layers with different numbers of neurons, while the code layer often contains a smaller number of neurons than those present in the input layer. Similar to the shallow autoencoder, DAE is commonly used to learn the advanced features of drugs/targets in drug repositioning [30, 49, 50], while the advanced features are fed into classifiers to make predictions.

Generative adversarial network (GAN) is based on a game theory that two neural networks compete with each other [104]. The two neural networks are the generator network and discriminator network, as shown in Fig. (**10b**). The

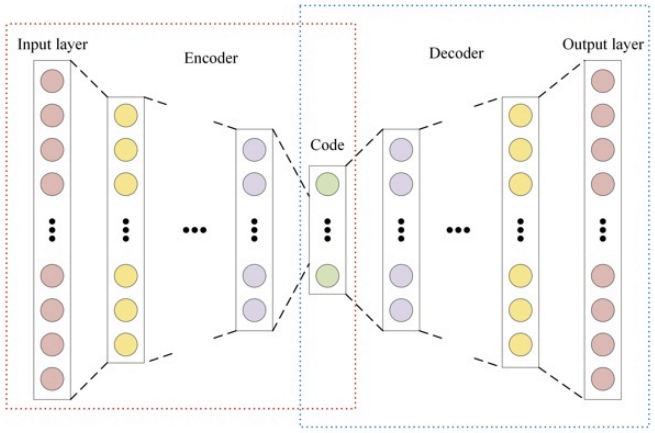


(a). Convolutional neural network

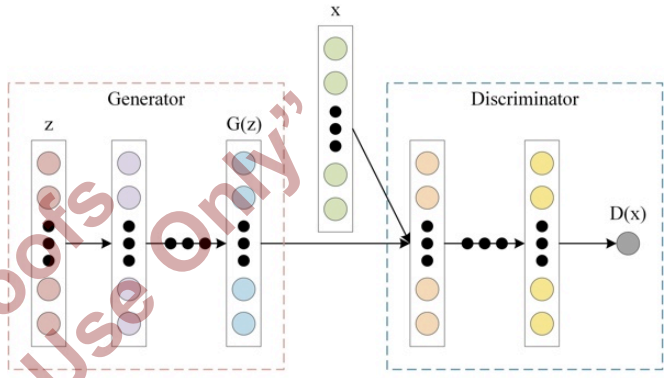


(b). Recurrent neural network

Fig. (9). The structures of convolutional neural network (**a**) and recurrent neural network (**b**).



(a). Deep autoencoder



(b). Generative adversarial network

Fig. (10). The structures of Deep autoencoder (**a**) and Generative adversarial network (**b**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

generator produces samples, and the discriminator aims to distinguish between the training samples and the samples from the generator [106]. Researchers employed the GAN models to distinguish the known DTIs and the unknown DTIs based on their feature vectors [107].

4. DRUG REPOSITIONING PROBLEMS

In the previous two sections, we have discussed the databases and ML/DL methods. In this section, we review some latest applications in drug repositioning. We divide the predictions of novel drugs into three types: drug-disease association (DDA) prediction, drug-drug interaction (DDI) prediction, and drug-target interaction (DTI) prediction. The DDA prediction aims to find some novel drugs directly, based on multiple types of drug features and disease features, such as drug structures, drug side effects, disease phenotypes, and disease genes. The second type is to identify some drugs which have interactions with existing drugs, as they may provide potential treatments for the same disease. The third type aims to identify some novel DTIs. Mostly, a drug target is a protein, which has essential functions in disease pathways.

4.1. Evaluation Metrics

In related studies in drug repositioning, the predictions of DDA, DDI, and DTI are often treated as binary classifications. Various evaluation metrics are used to measure the prediction performance.

Precision, recall and F1 score are commonly used to measure the prediction performance. Based on the four basic metrics of true positive (TP), false positive (FP), false negative (FN), and true negative (TN), as shown in Table **4**, precision is defined as TP/(TP+FP), recall is defined as TP/(TP+FN), and the F1 score is as follows:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

A receiver operating characteristic (ROC) curve is created by plotting the TP rate against the FP rate in various thresholds. Similarly, the precision-recall (PR) curve is created by plotting the precision against recall in various thresholds. Furthermore, the area under the ROC curve (AUCROC) and area under the PR curve (AUCPR) are used to measure the prediction performance, which is the area under the corresponding curve.

4.2. Drug-Disease Association Predictions

When handling disease and drug data together, the drug-drug similarities and disease-disease similarities are utilized by many drug repositioning methods. A small molecule drug may have many features, including side effects, chemical structures, target sequences, and related genes [108, 109].

Table 4. The confusion table.

-		Predicted Condition	
		Predicted Positive	**Predicted Negative**
Actual Condition	Actual positive	True Positive (TP)	False Negative (FN)
	Actual Negative	False Positive (FP)	True Negative (TN)

Different features may emply different methods for calculating similarity, as listed in Table **2**. A disease also has some features and similarity calculating methods [110, 111].

Based on the similarities, some machine learning methods are applied to predict potential drug-disease associations (DDAs), such as random walk [33, 34], SVM [43], and RF [23, 24]. Luo *et al*. applied one type of similarity for each instance and a random walk algorithm to identify new indications for existing drugs [33, 34]. In an earlier research work [33], the drug-drug chemical structure similarity and disease-disease phenotype similarity have been proposed to construct a drug similarity network and a disease similarity network. The two networks were connected by known DDAs and formed a heterogeneous network. A bi-random walk algorithm was applied to the heterogeneous network, while one random walk was in the drug network and another in the disease network. Each random walk produced a value, and the average value denoted the probability of the drug-disease association. In another study [34], the heterogeneous network contained three parts: drug network, disease network, and target network. A random walk with restart (RWR) was applied to the heterogeneous network and produced a probability vector, which contained the probability scores of all drugs associated with a given disease.

The drug-disease association prediction problem is often formulated as a classification problem. Lee-Yoon *et al*. constructed an RF model to predict potential DDAs via genes [23]. The genes were utilized to connect drug target genes and disease genes. Then the drug-disease pairs were represented to gene paths, which have been proposed to train an RF model. The known DDAs were assigned as positive samples, and the unknown ones as negative samples. Zhou *et al*. generated a drug-disease heterogeneous network and utilized an RF model to make a prediction [24].

Besides single similarity for drugs and diseases, multiple similarities can be concatenated together to increase the prediction accuracy. Kim *et al*. utilized four types of drug-drug similarity and three types of disease-disease similarity in their work [43]. Furthermore, 1,330 known DDAs were utilized as the basic instances. For a drug-disease association that needs to be predicted, the drug in it may have similarities with the drugs in all known DDAs, and the disease in it may have similarities with the diseases in all known DDAs. One type of drug similarity and one type of disease similarity are used to construct a classification feature. Twelve types of feature integrations are generated. Finally, an SVM model is constructed, and 10-fold cross-validation is applied to evaluate this model.

Besides basic ML methods, some DL methods are utilized to make the prediction. Liu *et al*. constructed a drug-disease heterogeneous network and applied a DNN model to predict potential DDAs [25]. An adjacent matrix has been constructed, while each row or column was treated as the feature vector of an instance. The two feature vectors of a drug-disease pair were integrated and fed into a DNN model, and a probability score has been generated. The proposed deep learning method achieves higher scores in multiple measurements than some ML approaches, including logistic regression, SVM, and RF.

Jarade *et al*. proposed a DNN model [53] and a collective variational autoencoder (cVAE) model [48] to predict DDAs. In their work, several drug similarities and disease similarities were filtered and integrated. The integrated feature vectors were fed into either a DNN model or a cVAE model to finish the prediction. The two models performed better than some machine learning approaches in terms of both AUCROC and AUCPR.

Zeng *et al*. proposed a multi-modal deep autoencoder (MDA) model to extract low-dimensional features from multiple networks and a cVAE model to predict potential DDAs [66]. A co-occurrence matrix was generated *via* a random walk on the heterogeneous network. Then the co-occurrence matrix has been transformed into a positive pointwise mutual information (PPMI) matrix [112], which was utilized as the input data of MDA [102]. The middle layer of the MDA informative feature was made part of the input of the cVAE model. Other parts of input data were the known DDAs. The probability score has been generated to reflect the potentiality of the drug-disease pairs.

Based on multiple features and similarities, Jiang *et al*. proposed an autoencoder model [26] and a CNN model [27] to predict potential associations. In the study in which the autoencoder model has been proposed [26], for a given drug-disease association, the drug chemical structure fingerprint, drug Gaussian interaction profile (GIP) kernel similarity, disease GIP kernel similarity [113], and disease MeSH term similarity were concatenated [114] and fed into an autoencoder. After dimensionality reduction, an RF classifier has been applied to finish the prediction. In the latter study [27], the autoencoder has been replaced by a CNN model, and the RF utilized as a classifier.

CNN is another commonly used DL model in drug repositioning. It can effectively extract features from different types of raw data. Li *et al*. proposed a CNN model to conduct a binary classification of DDAs [11]. The drug features were based on the simplified molecular-input line-entry sys-

tem (SMILES) [115] with a dimensionality of 881. The disease features were retrieved from the human symptoms-disease network [116], and its dimensionality was 322. An 881×322 matrix has been constructed and mapped to a gray-scale image. A CNN model has been then applied to extract feature vectors from the image and generate the prediction results.

Graph neural network (GNN) [117] has several subtypes, including graph convolution network (GCN) and graph auto-encoder (GAE). Wang *et al*. proposed a GNN based method to predict potential DDAs [118]. A drug-disease association network has been constructed from known associations. Then a GNN model was applied to exploit the high-order features in the network. Yu *et al*. came up with a layer attention GCN (LAGCN) model to predict DDAs after the construction of a drug-disease heterogeneous network [119]. In the embedding process of LAGCN, each layer involved a weight parameter to adjust the contribution of different layers. The parameters were auto-learned by NN.

In a previous research on DDAs, most of their features have been different; for instance, the drugs' chemical structures and the diseases' phenotype ontologies. However, both of them had associations with genes, which could be measured in microarray platforms. Focusing on the expression values of genes under different drugs in different cell lines can reveal the DDAs directly. In this way, a set of drug perturbation profiles can be downloaded from the CMap and LINCS databases, and the disease profiles can be downloaded from the GEO database, as listed in Table **1**.

Wang *et al*. applied a *k*-means algorithm to cluster the disease profiles into several groups to represent the cancer subtypes [99]. Each group was utilized to identify a list of disease genes. The disease-gene signatures have been based on the weighted frequencies of genes in the lists, which were mapped with the drug perturbation profiles in the CMap database [120, 121]. The connection score of a disease signature and a drug profile represents their possible association, while a negative number indicates that the drug may have potential treatments for the disease. In comparison with the methods without the *k*-means algorithm, the proposed framework achieves better prediction accuracy in several types of cancers. Zhao *et al*. used the drug profiles in CMap to train five machine learning classifiers. Based on the drug indications extracted from ATC and MEDI-HPS [122], the positive and negative drug labels have been generated. The authors focused their study on three types of diseases and predicted several drugs having literature evidence.

4.3. Drug-Drug Association Predictions

Unlike the drug-disease associations, the drug-drug interactions (DDIs) have the same feature types connecting them together, such as chemical structures, targets, enzymes, pathways, transports, indications, and side effects. There are many types of DDIs, which reflect the connections between the two drugs, such as the bioavailability/metabolism/serum concentration/therapeutic efficacy of drug *a* can be decreased/increased by drug *b*. Therefore, identifying the types of DDIs can help study the drug repositioning potentiality of a drug combination. Additionally, for a single drug, based on the "guilt-by-association" principle, the high similarities with other drugs may reflect their treatment similarity. Those two parts are the main field of DDI prediction for drug repositioning.

Ferdousi *et al*. employed 12 binary features to analyze DDIs [123]. The features were integrated, and the pair similarities have been calculated. For the known DDIs, a preprocessing step is added to delete the DDI whose two drugs have no common biological item or have an empty common feature vector. Among the remaining known DDIs, the minimum positive similarity value is set to be the threshold, which is utilized to determine whether an unknown drug-drug pair has the potential to be a DDI.

Yan *et al*. only calculated the similarities of known DDIs and applied a regularized least squares (RLS) classifier to finish the prediction [32, 124]. In an earlier work [32], eight types of drug features were integrated and the total dimensionality of the drug vector was kept as 21,351. Then the similarity of a drug-drug pair was calculated. Based on the known DDIs and similarities, the initial score of an unknown DDI was generated through the KNN method. The drug interaction vector consists of initial scores between it and all other drugs. The GIP kernel similarity matrix is based on the drug-drug interaction vectors. Finally, an RLS classifier is employed to predict potential DDIs based on the matrix. In another research work [124], the GIP similarity has been applied to the adjacent matrix directly, without the initial score procedure. Then the GIP similarity and drug feature cosine similarity have been integrated and averaged to construct the similarity matrix.

In many classification methods, the known DDIs are treated as positive samples, and unknown DDIs as negative samples. Some researchers identify reliable negative samples (RNS) from unknown DDIs. Bi *et al*. calculated an average distance between an unknown DDI and all known DDIs, while only the unknown DDIs with large distances were identified as RNS [125]. The residual unknowns were treated as unlabeled samples. The samples with three types of labels have been utilized for training an extreme learning machine (ELM) [126] and predict the potential DDIs. Zheng *et al*. applied an SVM to identify RNSs and another SVM to predict DDIs [44]. Its performance is better than Bi's method based on the measurement of recall and F-score.

In many studies, researchers prefer to use multiple types of similarities without any distinction. Rohani *et al*. added a filter procedure and employed a neural network model to predict potential DDIs [10]. They first selected several types of similarities with the most information and least redundancy [127], then a nonlinear method has been applied to integrate the selected similarity matrices. Each drug has a feature vector in the integrated matrix [128]. A neural network model integrates two drug feature vectors, and the output is a probability value for potential DDI.

DDIs can be used to construct a DDI network, where the known DDIs are the edges in the network. Then the DDI

prediction problem is transformed into the prediction of missing links in the network. Zhou *et al*. employed a Markov clustering algorithm to identify drug groups from the network, and most of the groups have been significantly correlated with certain functions [129]. Munir *et al*. applied the k-means algorithm to generate 12 clusters of drugs and construct 12 DDI networks [130]. All the drugs have been used for the treatment of epidermal growth factor receptor (EGFR) mutations in various cancers. The drugs that connected with the nodes with the largest centrality values in each network were selected and combined to construct a final DDI network. Then the same procedure has been applied to identify the final drugs with potential interactions. The predicted DDIs have been verified by molecular docking results.

Kastrin *et al*. integrated DDI networks with feature similarities to predict potential DDIs [22]. Their five networks have been found to be based on five databases. Five machine learning algorithms, including decision tree, KNN, SVM, RF, and gradient boosting machine (GBM), have been applied to finish the prediction based on topological features of the networks and semantic features.

Zhang *et al*. integrated 14 types of similarities to make the DDI prediction [35]. Eight of them were based on drug features, such as chemical structure, targets, and pathways. Six of those have been based on the DDI network, which has been constructed from the known DDIs. A random walk method was applied to the DDI network with each of the similarity matrices. All the predictions were combined through an ensemble learning procedure [131] to generate an improved final prediction result.

Similar to the drug-disease association predictions, DL methods are utilized to predict potential DDIs. Zhang *et al*. applied multi-modal deep auto-encoders to generate low-dimensional feature vectors of drug pairs and predicted potential DDIs via RF classifier [28]. Ryu *et al*. employed a DNN model to predict potential DDI types [54]. Shukla *et al*. proposed a modified DNN model to make the prediction [55]. In their model, a few CNN and RNN hidden layers were added to process the drug features; the prediction accuracy of their model has been found to be better than either CNN models or RNN models. Lee *et al*. collected three types of data, including drug structures, target genes, and GO terms [49]. For a given drug pair, three types of feature vectors have been selected. The same types have been integrated and fed into an autoencoder. The three code layers of the three autoencoders have been integrated again and fed into a DNN model, which has been used to predict DDI types. Deng *et al*. utilized four types of similarities and constructed four similarity matrices [56]. The similarity matrices were fed into a DNN model, and the output was the DDI events, which have been used to describe the DDI relationships. Feng *et al*. proposed a GCN model to extract the network structure features of drugs from the DDI network and predict DDIs [57]. A 2-layer GCN was utilized to obtain drug features and produce a feature vector matrix. Two drug vectors were integrated and fed into a DNN model, which was used to deduce the potential DDIs.

The previous studies have involved drug-drug pairs, as in some conditions, a combination of more than two drugs may provide potential treatments. Peng *et al*. proposed a novel model to predict the reactions of drug combinations [132]. In the first process, the dimensionalities of drug features were reduced through a neural network model. The new drug vectors were integrated via three approaches: max pooling, mean pooling, and self-attention. The embedding vectors were fed into a second neural network model, and the output value was used to predict the potential reactions of the drug combination.

Some researchers have been observed to add more entities to the DDI network and construct a new knowledge graph to reflect the new associations. Lin *et al*. utilized drugs, targets, genes, transporters, and enzymes to build a knowledge graph [133]. The drug feature vector of a drug-drug pair was encoded by a 2-layer GNN model. Then the output values were used to predict whether the drug-drug pair has potential interactions.

In many methods, two drugs in a DDI are treated separately. Song *et al*. implemented a different idea to make the prediction [45]. In their method, the drug-drug pairs were treated as instances. The dug pair similarities have been calculated based on the drug similarities as follows:

$$S\left(\left(d_1,d_2\right),\left(d_3,d_4\right)\right) = \max\left(S\left(d_1,d_3\right)\times\left(d_2,d_4\right),\left(d_1,d_4\right)\times\left(d_2,d_3\right)\right) \quad (5)$$

Where, $S(i,j)$ is the similarity between two instances i and j, and (d_1,d_2) is a drug-drug pair. An SVM model has been proposed to make the subsequent prediction. A DDI's feature is determined by its similarities with other DDIs. Like the training strategy in other methods, 10-fold cross-validation is applied to evaluate the SVM model. In the results, some DDIs with literature evidence have been predicted, which are not listed in the referenced databases.

Cytochrome P450 enzymes are essential for the metabolism of many medications [134], which are the main reasons for many DDIs. A drug can be a substrate, inhibitor, or inducer of CYP450, which may affect the metabolite of other drugs. Hunta *et al*. predicted potential DDIs via their enzyme actions [46]. Different from other features of drugs, the features in Hunta's study included enzymes and enzyme action types. Machine learning algorithms, such as NN and SVM, were trained and used to predict the potential DDI.

4.4. Drug-Target Association Predictions

A target is a molecule that has a proven association with a particular disease [71]. It is usually a protein. In recent years, many databases and tools have been constructed to reveal interactions between diseases and genes or proteins, which has helped researchers predict potential drugs through drug-target interactions (DTI).

The decision tree, RF, and SVM are commonly used classification algorithms in machine learning that many researchers employ in drug repositioning. Wang *et al*. applied

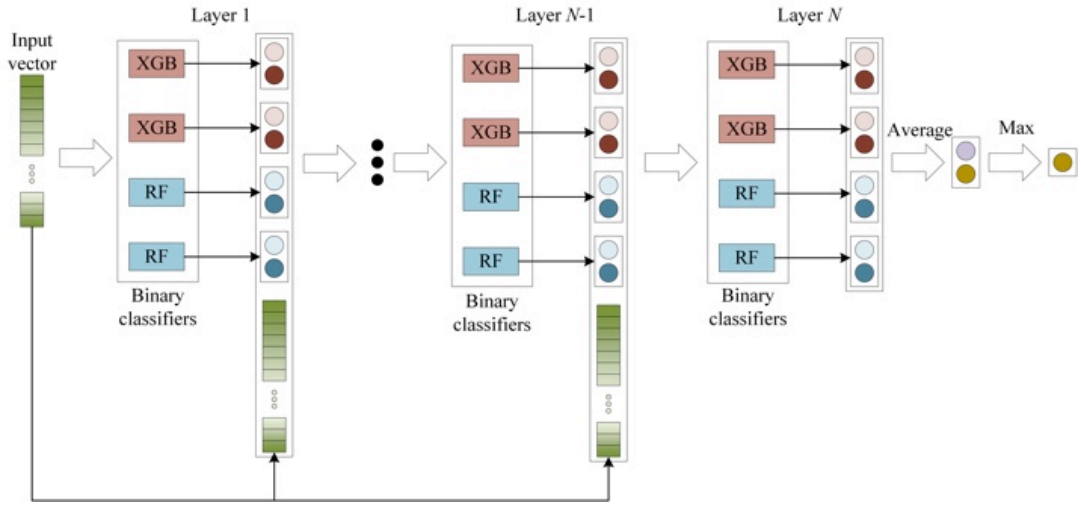


Fig. (11). The structure of CDF. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

an RF approach to predict DTIs [29]. In their method, the protein-ligand connection was described by four components: protein sequence, binding pocket, ligand structure, and intermolecular interaction. In general, the total number of features is several thousands. A PCA procedure has been employed to reduce the dimensionality of features before the RF model. The number of final features became less than a few hundred. After training, their method exhibited good results in predicting DTIs.

Similar to the drug-disease heterogeneous network, researchers have constructed a drug-target network to predict potential DTIs. The drug-drug similarities and target-target similarities are calculated from various features, and the known DTIs are downloaded from public databases. Based on the heterogeneous network and similarity matrices, Zeng *et al*. generated feature vectors of drugs and targets separately [51]. A deep forest (DF) classifier has been applied to predict potential DTIs from the feature vectors.

Chu *et al*. utilized a cascade deep forest (CDF) model to predict potential DTIs [135]. A few steps were utilized to generate the features, which have been inputted into the model. Six types of similarities have been used to construct the drug-target heterogeneous networks. The networks then have been merged by a network fusion method [128]. In Chu's work, they used the path nodes between the drug and the target to form the input vector [127]. The path node has been either a different drug or a different target, restricted to be the five nearest neighbors of the initial drug and target. As a result, the new form of input vector may be drug-drug-target, drug-drug-drug-target, or four other forms. After being fed into the CDF model [136], as shown in Fig. (**11**), a final prediction has been made from the output. In each layer of CDF, the number of binary classifiers was found to be varied.

Lin *et al*. utilized support vector regression (SVR) to build a model to predict potential DTIs [137]. In their study, the SVR has been applied to generate the binding strength of drug-protein pairs. A protein similarity network has been constructed, where the similarities were found to be based on the binding strength. The edge betweenness centrality has been used to predict shared drugs between proteins, which may constitute the potential DTIs.

Zong *et al*. utilized a DeepWalk method [36], which is a deep model of random walk, to predict DTIs from a network model [138]. The known DDAs, DDIs, and DTIs were downloaded and used to construct a drug-target-disease network. The similarity between two instances has been calculated by DeepWalk based on the known edges. After generating the similarities, two approaches have been proposed to predict potential DTIs, which are drug-based and target-based similarity inferences [139].

Many researchers apply the drug-target heterogeneous network to identify their feature vectors. The dimensionality of each vector is the sum of drug features and disease features. Manoochehri *et al*. proposed a different approach to generate the feature vectors from the drug-target network [12, 13]. For a drug-target pair, the sub-graph is constructed based on their neighbors in the network and themselves, which means that different interaction leads to the formation of different sub-graphs. An adjacent matrix is identified based on the sub-graph rather than the whole drug-target network. Therefore, the feature vectors also have different dimensionalities. After the features are fed into an NN model, a prediction is made. When training the model, the known DTIs produce known sub-graphs for positive samples, and the negative samples are not selected randomly but built under certain principles [140]. After training, the proposed method achieves higher performance than the baseline methods in terms of AUCROC and AUCPR.

Although the basic ML methods achieve satisfying prediction performance, the DL methods work better in many cases. Wen *et al*. proposed the first deep learning method (DeepDTI) for predicting DTIs [52]. The drug substructure fingerprints were identified as the drug feature vectors, and the target protein sequences were target vectors. The DeepDTI invovles a deep-belief network (DBN), which is made by stacking restricted Boltzmann machines (RBMs). In various measurements of predictions, the DeepDTI method

achieves better performance than other ML methods, including RF, decision tree, and naive Bayesian.

When applying a DNN model to predict DTIs from drugs and target feature vectors, some basic ML and DL algorithms are also utilized to generate satisfied feature vectors, such as linear classification [141, 142], random walk with restart (RWR) [37-40], autoencoder [19, 30, 41, 42, 50], *etc*. Parvizi *et al*. utilized the random walk with restart (RWR) algorithm and skip-gram neural network to generate the feature vectors of drugs and targets [40]. In their method, the drug-target heterogeneous network was replaced by two networks: drug-related network and protein-related network. The drug-related network consists of DDIs and DDAs, while the protein-related network contains protein-protein interactions (PPIs) and protein-disease associations.

Peng *et al*. proposed a similar approach to constructing a drug-related network and a protein-related network [61]. Besides the known interactions and associations, the drug-drug similarities and protein-protein similarities were added in the networks. After integration of the two feature vectors, a deep autoencoder has been applied to produce the low dimensional features, which have been fed into a CNN model. The prediction performances in terms of AUCROC and AUCPR were increased by adding the similarities, which resulted in higher prediction performance compared to other ML methods.

CNN is a commonly used model for deep learning. It can be applied to either make the prediction or produce satisfied feature vectors of drugs and proteins. Hu *et al*. utilized a CNN model to predict DTIs [14]. The drug chemical structure vectors from PaDEL-descriptor [143] and the target amino acid physicochemical property vectors from AAindex [144] have been proposed to identify the input matrix of the CNN method. The combination of drug vector and target vector was randomly selected, and the combinations of known DTIs were treated as positive samples, while others as negative samples. With the 10-fold cross-validation, the prediction performances were found to be much better than the state-of-the-art methods.

Monteiro *et al*. used CNN models to identify the feature vectors and apply a DNN model to predict DTIs [15]. After generating a drug SMILES vector and a target sequence vector from databases, two CNN models have been proposed to process the two types of feature vectors and produce two novel vectors. The two vectors were integrated and fed into a fully connected DNN model. Finally, a prediction of DTI has been made. Compared with the method without the CNN pre-processing and the CNN-RF/SVM models, the CNN-DNN architecture yielded improved results in the correct classification of both positive and negative interactions.

Similarly, Öztürk *et al*. [58] and Zhao *et al*. [59] applied CNN and DNN models to generate the feature vectors and predict potential DTIs. In Öztürk's method, the drug SMILES features and protein sequence features were processed by two CNNs separately. The two feature vectors of a drug-target pair were generated, integrated, and fed into a DNN model to make a prediction. In Zhao's method, the commonly utilized drug-target heterogeneous network was transformed into a drug-target pair (DTP) network. Different from the heterogeneous network, the nodes in the DTP network were the drug-target pairs. The number of pairs in the DTP network was $n \times m$, with n being the number of drugs and m the number of targets. A GCN model has been processed to extract features from the adjacent matrix of the network. The new features were fed into a DNN model, and the prediction has been made.

Huang *et al*. proposed a deep learning library to predict DTIs [16]. In their library, only the drug SMILES vectors and protein amino acid sequence vectors were utilized. Those two vectors were transformed into two new feature vectors through utilizing 15 approaches, such as CNN and RNN. Then the two feature vectors were integrated and fed into a multi-layer perceptron to generate the prediction of the drug-target pair.

Lee *et al*. proposed an integrated model to make the prediction [17]. In their method, a convolution layer has been applied to process the target sequences, and a fully connected layer was used to process drug fingerprints. Then two vectors were integrated and fed into a CNN model. They compared their method with DeepDTI [52], which has been discussed previously. The DeepConv-DTI achieved higher accuracy and F1 score.

Similar to the DDA and DDI predictions, GNN is widely used for predicting DTIs. Jiang *et al*. utilized a GNN model to identify the feature vectors, and then an NN model has been applied to predict DTIs [18]. The drug's chemical structure was proposed to construct a molecular graph. The nodes were atoms, and the edges were bonds. The protein amino acid graph, based on the protein contact map, was produced by PconsC4 [145] based on the amino acid sequences. A new drug vector and target vector were identified by the GNN model, and their integration was fed into an NN model to make the prediction.

Lim *et al*. constructed a different graph based on the protein-ligand complex [146]. The structure information of protein and ligand atoms was embedded in two adjacent matrices, A^1 and A^2. A^1 contained covalent interactions only, and A^2 contained both covalent interactions and non-covalent intermolecular interactions. Two node feature vectors were generated from either A^1 or A^2. By subtracting the two feature vectors, their difference was fed into a GNN, and the prediction results have been generated.

In many studies, one drug vector is integrated with one target vector. It is crucial to determine which type of target feature is used to identify the integration. In Lee's research, three types of targets have been proposed to have close relationships with protein functions or drug mechanism of actions (MoAs) [60]. One drug vector, based on differentially expressed genes from the LINCS database [147], was integrated with all three target vectors, including gene knockdown expression profiles (GEPs) from the LINCS database, protein-protein interaction (PPI) network from String database [148], and pathway memberships from MSigDB [149]. After integration, the new vector was fed into a DNN model.

The concatenation of three types of target vectors exhibited better performance in terms of AORUC than any single type.

Agyeman *et al*. proposed integrated views predictive GAN (IVPGAN) to predict potential DTIs [107]. The model contains two main parts, generator and discriminator. The input data of the generator is the integrated vector of drug graph representation, drug SMILES string, and target sequence. The output of the generator, which reflects the binding strength, is combined with the ground truth and fed into the discriminator. Like other DL methods, the authors utilized a 5-fold CV to evaluate the IVPGAN model, and the prediction performance was found to be higher than the parametric models in most of the datasets.

In the previous studies, many methods have been used to integrate the feature vectors of drugs and targets directly, but have failed to learn the low-dimensional features. Autoencoder is an excellent unsupervised approach to reduce dimensionality with excellence. Wang *et al*. applied a stacked autoencoder to identify protein features from sequence information [30]. An RF classifier has been utilized after the integration of protein feature vectors and drug structure vectors. Sun *et al*. proposed a convolutional autoencoder and GAN-based method to predict DTIs [41]. After constructing a drug-target heterogeneous network, the adjacent matrix has been fed into a convolutional autoencoder, and a novel feature matrix with lower dimensionality has been generated. It was assumed that the new feature vector of a drug or target would obey the Gaussian distribution. After the discriminator step, a prediction of DTI has been made. In the evaluation, the proposed method achieved better performance than some DTI prediction methods, including DTINet by Luo *et al*. [37], Lee's method [38], and DTIGBDT proposed by Xuan *et al*. [39]. The RWR algorithm has been applied to capture topological information in the networks of these models.

Torng *et al*. applied a graph autoencoder (GAE) to extract a representation of protein pocket features [19]. Before the final classifier, a fully connected layer was added, taking the joint vector of protein and drug as input and producing a low-dimensional hidden layer as output. In evaluation, the proposed method outperformed several structure-based and ligand-based methods in AUCROC scores.

Wang *et al*. utilized a multi-modal deep autoencoder (MDA) to produce protein and drug feature vectors based on several similarities [42]. Each type of similarity involved a corresponding network. In each network, the RWR method and PPMI were applied to calculate the topological similarity of drugs and proteins. Then the global structure information has been generated. Two MDAs were applied to integrate multiple similarity measures of drugs and targets and learn their low-dimensional feature matrices. The two features of a drug and a target were merged and fed into a DNN to make a prediction.

Since last year, COVID-19 has spread havoc all over the world. Many researchers have focused on either vaccines or medications to help stop the pandemic. With the development of SARS-CoV-2's core proteins, Beck *et al*. used natural language processing (NLP) to identify potential DTIs [150]. In NLP, the molecule sequence is analogous to a language. More than 1 million drugs are used to train the models, and several antiviral drugs have been proposed that have potential interactions with SARS-CoV-2 proteins. Remdesivir, which has been reported to be an effective medication for COVID-19 *in vitro* [151], is among the prediction results.

5. DISCUSSION

In the former sections, we have reviewed some latest studies employing basic ML and DL to predict novel drugs. Various methods have been used to predict the potential DDAs, DDIs, and DTIs. Those predictions help to find novel treatments for existing drugs. In some cases, researchers have also identified some potential drugs for specific diseases by using the proposed methods. However, there are still some limitations.

A general issue is related to the feature types in databases. As shown in Table **1**, there are a large number of databases that store the drug-, disease- and target-centric information. Some databases may focus on a single feature type for each category (drug, disease, or target), while others may be comprehensive. The credibility of different drug features makes a big difference. A similar statement can be made for disease features and target features. In many studies, only one feature type for each category is applied. Although the drug chemical structure, disease phenotype, and protein amino acid sequence are widely used, other types should not be ignored. In previous studies [35, 49, 51], the feature vectors have been identified from multiple feature types. However, it is still important to select several reliable feature types. In a study [60], Lee *et al*. proposed three types of target features closely related to DTIs. The selection of different feature types is now attracting attention.

When using multiple feature types, a second issue is how to effectively integrate them together. Researchers use many different strategies to perform the integration. In studies conducted earlier [29, 32], the multiple feature vectors for the same category (drug, disease, or target) were concatenated directly, without any additional processing. The ML models were then constructed based on the integrated data. In other studies [10, 22, 135], several approaches were used to integrate the feature vectors or similarities, such as the average similarity of multiple types. In a research work carried out previously [35], the authors constructed 29 models based on the multiple feature types, and then the results were merged together to identify the prediction. Although all of the different strategies generate satisfied predictions, further ensemble methods need to be proposed.

A third issue which needs to be improved is the identification of negative samples. A large number of applications use basic ML and DL models to classify DDAs (or DDIs, DTIs). In many studies, the negative samples have been randomly selected from the unknown associations. To improve the accuracy of samples, a few strategies have been proposed to identify reliable negative samples (RNS). In a research

work [152], the authors first used the known and unknown associations to construct a classifier, then they employed this classifier to classify the unknown associations. The authors classified negative samples and identified them as RNS. In a few studies conducted earlier [44, 125, 153], KNN, RWR, and SVM were applied to extract RNS. Besides calculating distances and similarities, more reliable strategies are needed.

The fourth issue is related to the use of ML and DL methods. These methods are just like black boxes, which make the models lack interpretability. Compared with DL models, some basic ML models are more interpretable, such as decision tree and logistic regression. Meanwhile, compared to basic ML models, DL models achieve better performance in predicting potential DDAs, DDIs, and DTIs. Therefore, more interpretable ML and DL models are essential in many application domains, especially in human healthcare-related fields [154], where drug repositioning is applied. To achieve this goal, the improvements of basic ML and DL models with interpretability are necessary.

CONCLUSION

In this study, we have reviewed some latest studies predicting novel treatments for existing drugs. The widely used databases and pre-processing steps have been introduced. The six data types in those databases, including drug features, disease features, target features, DDAs, DDIs, and DTIs, have been taken into consideration. We have then discussed commonly used basic ML and DL methods, and their applications to the predictions of DDAs, DDIs, and DTIs. To address the limitations of existing methods, we suggest future works to be directed towards features, samples, and methods, which could benefit the research community in drug repositioning.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

This work has been supported in part by the Natural Science and Engineering Research Council of Canada (NSERC), the China Scholarship Council (CSC) and by the National Natural Science Foundation of China under Grant Nos. (61772552 and 61428209) .

CONFLICT OF INTEREST

Dr. Fang-Xiang Wu is the Section Editor of the journal Current Bioinformatics.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Emmert-Streib F, Tripathi S, Simoes RD, Hawwa AF, Dehmer M. The human disease network: Opportunities for classification, diagnosis, and prediction of disorders and disease genes. Syst Biomed 2013; 1(1): 20-8. http://dx.doi.org/10.4161/sysb.22816

[2] Matthews H, Hanison J, Nirmalan N. "Omics"-informed drug and biomarker discovery: Opportunities, challenges and future perspectives. Proteomes 2016; 4(3): 28. http://dx.doi.org/10.3390/proteomes4030028 PMID: 28248238

[3] Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: Progress, challenges and recommendations. Nat Rev Drug Discov 2019; 18(1): 41-58. http://dx.doi.org/10.1038/nrd.2018.168 PMID: 30310233

[4] Jin G, Wong ST. Toward better drug repositioning: Prioritizing and integrating existing methods into efficient pipelines. Drug Discov Today 2014; 19(5): 637-44. http://dx.doi.org/10.1016/j.drudis.2013.11.005 PMID: 24239728

[5] Swinney DC, Anthony J. How were new medicines discovered? Nat Rev Drug Discov 2011; 10(7): 507-19. http://dx.doi.org/10.1038/nrd3480 PMID: 21701501

[6] Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: From data to therapeutics. Clin Pharmacol Ther 2013; 93(4): 335-41. http://dx.doi.org/10.1038/clpt.2013.1 PMID: 23443757

[7] Szabo M, Svensson Akusjärvi S, Saxena A, Liu J, Chandrasekar G, Kitambi SS. Cell and small animal models for phenotypic drug discovery. Drug Des Devel Ther 2017; 11: 1957-67. http://dx.doi.org/10.2147/DDDT.S129447 PMID: 28721015

[8] Santos R, Ursu O, Gaulton A, *et al.* A comprehensive map of molecular drug targets. Nat Rev Drug Discov 2017; 16(1): 19-34. http://dx.doi.org/10.1038/nrd.2016.230 PMID: 27910877

[9] Turing AM. Computing machinery and intelligence. 2009. http://dx.doi.org/10.1007/978-1-4020-6710-5_3

[10] Rohani N, Eslahchi C. Drug-drug interaction predicting by neural network using integrated similarity. Sci Rep 2019; 9(1): 13645. http://dx.doi.org/10.1038/s41598-019-50121-3 PMID: 31541145

[11] Li Z, Huang Q, Chen X, *et al.* Identification of drug-disease associations using information of molecular structures and clinical symptoms *via* deep convolutional neural network. Front Chem 2020; 7: 924. http://dx.doi.org/10.3389/fchem.2019.00924 PMID: 31998700

[12] Manoochehri HE, Kadiyala SS, Nourani M. Predicting drug-target interactions using weisfeiler-lehman neural network. Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2019 May 19-22; Chicago, IL, USA, 2019. http://dx.doi.org/10.1109/BHI.2019.8834572

[13] Eslami Manoochehri H, Nourani M. Drug-target interaction prediction using semi-bipartite graph model and deep learning. BMC Bioinformatics 2020; 21(Suppl. 4): 248. http://dx.doi.org/10.1186/s12859-020-3518-6 PMID: 32631230

[14] Hu S, Zhang C, Chen P, Gu P, Zhang J, Wang B. Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks. BMC Bioinformatics 2019; 20(25)(Suppl. 25): 689. http://dx.doi.org/10.1186/s12859-019-3263-x PMID: 31874614

[15] Monteiro NR, Ribeiro B, Arrais J. Drug-target interaction prediction: End-to-end deep learning approach. IEEE/ACM Trans Comput Biol Bioinform 2020; 2020: 1-12. http://dx.doi.org/10.1109/TCBB.2020.2977335

[16] Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: A deep learning library for drug-target interaction prediction. Bioinformatics 2021; 36(22-23): 5545-7. http://dx.doi.org/10.1093/bioinformatics/btaa1005 PMID: 33275143

[17] Lee I, Keum J, Nam H. DeepConv-DTI: Prediction of drug-target interactions *via* deep learning with convolution on protein sequences. PLOS Comput Biol 2019; 15(6): e1007129. http://dx.doi.org/10.1371/journal.pcbi.1007129 PMID: 31199797

[18] Jiang M, Li Z, Zhang S, *et al.* Drug–target affinity prediction using graph neural network and contact maps. RSC Adv 2020; 10(35): 20701-12.

http://dx.doi.org/10.1039/D0RA02297G

[19] Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions. J Chem Inf Model 2019; 59(10): 4131-49. http://dx.doi.org/10.1021/acs.jcim.9b00628 PMID: 31580672

[20] Wang Y, Deng G, Zeng N, Song X, Zhuang Y. Drug-disease association prediction based on neighborhood information aggregation in neural networks. IEEE Access 7: 50581-7. http://dx.doi.org/10.1109/ACCESS.2019.2907522

[21] Oh M, Ahn J, Yoon Y. A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. PLoS One 2014; 9(10): e111668. http://dx.doi.org/10.1371/journal.pone.0111668 PMID: 25356910

[22] Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. PLoS One 2018; 13(5): e0196865. http://dx.doi.org/10.1371/journal.pone.0196865 PMID: 29738537

[23] Lee T, Yoon Y. Drug repositioning using drug-disease vectors based on an integrated network. BMC Bioinformatics 2018; 19(1): 446. http://dx.doi.org/10.1186/s12859-018-2490-x PMID: 30463505

[24] Zhou R, Lu Z, Luo H, Xiang J, Zeng M, Li M. NEDD: A network embedding based method for predicting drug-disease associations. BMC Bioinformatics 2020; 21(Suppl. 13): 387. http://dx.doi.org/10.1186/s12859-020-03682-4 PMID: 32938396

[25] Liu H, Zhang W, Song Y, Deng L, Zhou S. HNet-DNN: Inferring new drug-disease associations with deep neural network based on heterogeneous network features. J Chem Inf Model 2020; 60(4): 2367-76. http://dx.doi.org/10.1021/acs.jcim.9b01008 PMID: 32118415

[26] Jiang HJ, Huang YA, You ZH. Predicting drug-disease associations *via* using Gaussian interaction profile and kernel-based autoencoder. BioMed Res Int 2019; 2019: 2426958. http://dx.doi.org/10.1155/2019/2426958 PMID: 31534955

[27] Jiang HJ, You ZH, Huang YA. Predicting drug-disease associations *via* sigmoid kernel-based convolutional neural networks. J Transl Med 2019; 17(1): 382. http://dx.doi.org/10.1186/s12967-019-2127-5 PMID: 31747915

[28] Zhang Y, Qiu Y, Cui Y, Liu S, Zhang W. Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning. Methods 2020; 179: 37-46. http://dx.doi.org/10.1016/j.ymeth.2020.05.007 PMID: 32497603

[29] Wang Y, Guo Y, Kuang Q, *et al.* A comparative study of family-specific protein-ligand complex affinity prediction based on random forest approach. J Comput Aided Mol Des 2015; 29(4): 349-60. http://dx.doi.org/10.1007/s10822-014-9827-y PMID: 25527073

[30] Wang L, You ZH, Chen X, *et al.* A computational-based method for predicting drug-target interactions by using stacked autoencoder deep neural network. J Comput Biol 2018; 25(3): 361-73. http://dx.doi.org/10.1089/cmb.2017.0135 PMID: 28891684

[31] Kuo B, Kang Y, Wu P, Huang ST, Huang Y. Discovering drug-drug and drug-disease interactions inducing acute kidney injury using deep rule forests. Proceedings of the IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI); 2020 Aug. 11-13; Las Vegas, NV, USA.

[32] Yan C, Duan G, Pan Y, Wu FX, Wang J. DDIGIP: predicting drug-drug interactions based on Gaussian interaction profile kernels. BMC Bioinformatics 2019; 20(Suppl. 15): 538. http://dx.doi.org/10.1186/s12859-019-3093-x PMID: 31874609

[33] Luo H, Wang J, Li M, *et al.* Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. Bioinformatics 2016; 32(17): 2664-71. http://dx.doi.org/10.1093/bioinformatics/btw228 PMID: 27153662

[34] Luo H, Wang J, Li M, *et al.* Computational drug repositioning with random walk on a heterogeneous network. IEEE/ACM Trans Comput Biol Bioinform 2018; 16(6): 1890-900.

[35] Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics 2017; 18(1): 18. http://dx.doi.org/10.1186/s12859-016-1415-9 PMID: 28056782

[36] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014 Aug. 24-27; NY, USA. http://dx.doi.org/10.1145/2623330.2623732

[37] Luo Y, Zhao X, Zhou J, *et al.* A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun 2017; 8(1): 573. http://dx.doi.org/10.1038/s41467-017-00680-8 PMID: 28924171

[38] Lee I, Nam H. Identification of drug-target interaction by a random walk with restart method on an interactome network. BMC Bioinformatics 2018; 19(Suppl. 8): 208. http://dx.doi.org/10.1186/s12859-018-2199-x PMID: 29897326

[39] Xuan P, Sun C, Zhang T, Ye Y, Shen T, Dong Y. Gradient boosting decision tree-based method for predicting interactions between target genes and drugs. Front Genet 2019; 10: 459. http://dx.doi.org/10.3389/fgene.2019.00459 PMID: 31214240

[40] Parvizi P, Azuaje F, Theodoratou E, Luz S. A network-based embedding method for drug-target interaction prediction. Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2020 July 20-24; QC, Canada. http://dx.doi.org/10.1109/EMBC44109.2020.9176165

[41] Sun C, Xuan P, Zhang T, Ye Y. Graph convolutional autoencoder and generative adversarial network-based method for predicting drug-target interactions. IEEE/ACM Trans Comput Biol Bioinform 2020; 2020: 1.

[42] Wang H, Wang J, Dong C, Lian Y, Liu D, Yan Z. A novel approach for drug-target interactions prediction based on multimodal deep autoencoder. Front Pharmacol 2020; 10: 1592. http://dx.doi.org/10.3389/fphar.2019.01592 PMID: 32047432

[43] Kim E, Choi AS, Nam H. Drug repositioning of herbal compounds *via* a machine-learning approach. BMC Bioinformatics 2019; 20(Suppl. 10): 247. http://dx.doi.org/10.1186/s12859-019-2811-8 PMID: 31138103

[44] Zheng Y, Peng H, Zhang X, Zhao Z, Gao X, Li J. DDI-PULearn: A positive-unlabeled learning method for large-scale prediction of drug-drug interactions. BMC Bioinformatics 2019; 20(Suppl. 19): 661. http://dx.doi.org/10.1186/s12859-019-3214-6 PMID: 31870276

[45] Song D, Chen Y, Min Q, *et al.* Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies. J Clin Pharm Ther 2019; 44(2): 268-75. http://dx.doi.org/10.1111/jcpt.12786 PMID: 30565313

[46] Hunta S, Aunsri N, Yooyativong T. Drug-drug interactions prediction from enzyme action crossing through machine learning approaches. Proceedings of the 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON); 2018 July 18-21; Chiang Rai, Thailand. http://dx.doi.org/10.1109/ECTICon.2015.7207126

[47] Zhang W, Yue X, Huang F, Liu R, Chen Y, Ruan C. Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. Methods 2018; 145: 51-9. http://dx.doi.org/10.1016/j.ymeth.2018.06.001 PMID: 29879508

[48] Jarada TN, Rokne JG, Alhajj R. SNF–CVAE: Computational method to predict drug-disease interactions using similarity network fusion and collective variational autoencoder. Knowl Base Syst 2021; 212: 106585. http://dx.doi.org/10.1016/j.knosys.2020.106585

[49] Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug-drug interaction effects. BMC Bioinformatics 2019; 20(1): 415. http://dx.doi.org/10.1186/s12859-019-3013-0 PMID: 31387547

[50] Zeng X, Zhu S, Lu W, *et al.* Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci (Camb) 2020; 11(7): 1775-97. http://dx.doi.org/10.1039/C9SC04336E PMID: 34123272

[51] Zeng X, Zhu S, Hou Y, *et al.* Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. Bioinformatics 2020; 36(9): 2805-12. http://dx.doi.org/10.1093/bioinformatics/btaa010 PMID: 31971579

[52] Wen M, Zhang Z, Niu S, *et al.* Deep-learning-based drug–target interaction prediction. J Proteome Res 2017; 16(4): 1401-9.

http://dx.doi.org/10.1021/acs.jproteome.6b00618 PMID: 28264154
[53] Jarada TN, Rokne JG, Alhajj R. SNF-NN: computational method to predict drug-disease interactions using similarity network fusion and neural networks. BMC Bioinformatics 2021; 22(1): 28. http://dx.doi.org/10.1186/s12859-020-03950-3 PMID: 33482713
[54] Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug-drug and drug-food interactions. Proc Natl Acad Sci USA 2018; 115(18): E4304-11. http://dx.doi.org/10.1073/pnas.1803294115 PMID: 29666228
[55] Kumar Shukla P, Kumar Shukla P, Sharma P, *et al.* Efficient prediction of drug-drug interaction using deep learning models. IET Syst Biol 2020; 14(4): 211-6. http://dx.doi.org/10.1049/iet-syb.2019.0116 PMID: 32737279
[56] Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug-drug interaction events. Bioinformatics 2020; 36(15): 4316-22. http://dx.doi.org/10.1093/bioinformatics/btaa501 PMID: 32407508
[57] Feng YH, Zhang SW, Shi JY. DPDDI: A deep predictor for drug-drug interactions. BMC Bioinformatics 2020; 21(1): 419. http://dx.doi.org/10.1186/s12859-020-03724-x PMID: 32972364
[58] Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics 2018; 34(17): i821-9. http://dx.doi.org/10.1093/bioinformatics/bty593 PMID: 30423097
[59] Zhao T, Hu Y, Valsdottir LR, Zang T, Peng J. Identifying drug-target interactions based on graph convolutional network and deep neural network. Brief Bioinform 2021; 22(2): 2141-50. http://dx.doi.org/10.1093/bib/bbaa044 PMID: 32367110
[60] Lee H, Kim W. Comparison of target features for predicting drug-target interactions by deep neural network based on large-scale drug-induced transcriptome data. Pharmaceutics 2019; 11(8): 377. http://dx.doi.org/10.3390/pharmaceutics11080377 PMID: 31382356
[61] Peng J, Li J, Shang X. A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network. BMC Bioinformatics 2020; 21(13)(Suppl. 13): 394. http://dx.doi.org/10.1186/s12859-020-03677-1 PMID: 32938374
[62] Xuan P, Ye Y, Zhang T, Zhao L, Sun C. Convolutional neural network and bidirectional long short-term memory-based method for predicting drug–disease associations. Cells 2019; 8(7): 705. http://dx.doi.org/10.3390/cells8070705 PMID: 31336774
[63] Xuan P, Cui H, Shen T, Sheng N, Zhang T. HeteroDualNet: A dual convolutional neural network with heterogeneous layers for drug-disease association prediction *via* Chou's five-step rule. Front Pharmacol 2019; 10: 1301. http://dx.doi.org/10.3389/fphar.2019.01301 PMID: 31780934
[64] Xuan P, Zhao L, Zhang T, Ye Y, Zhang Y. Inferring drug-related diseases based on convolutional neural network and gated recurrent unit. Molecules 2019; 24(15): 2712. http://dx.doi.org/10.3390/molecules24152712 PMID: 31349692
[65] Xuan P, Gao L, Sheng N, Zhang T, Nakaguchi T. Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations. IEEE J Biomed Health Inform 2021; 25(5): 1793-804. http://dx.doi.org/10.1109/JBHI.2020.3039502 PMID: 33216722
[66] Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: A network-based deep learning approach to *in silico* drug repositioning. Bioinformatics 2019; 35(24): 5191-8. http://dx.doi.org/10.1093/bioinformatics/btz418 PMID: 31116390
[67] Palleria C, Di Paolo A, Giofrè C, *et al.* Pharmacokinetic drug-drug interaction and their implication in clinical management. J Res Med Sci 2013; 18(7): 601-10. PMID: 24516494
[68] Altshuler D, Daly M, Kruglyak L. Guilt by association. Nat Genet 2000; 26(2): 135-7. http://dx.doi.org/10.1038/79839 PMID: 11017062
[69] Oliver S. Guilt-by-association goes global. Nature 2000; 403(6770): 601-3. http://dx.doi.org/10.1038/35001165 PMID: 10688178
[70] Hu SS, Chen P, Wang B, Li J. Protein binding hot spots prediction from sequence only by a new ensemble learning method. Amino Acids 2017; 49(10): 1773-85. http://dx.doi.org/10.1007/s00726-017-2474-6 PMID: 28766075
[71] Gashaw I, Ellinghaus P, Sommer A, Asadullah K. What makes a good drug target? Drug Discov Today 2011; 16(23-24): 1037-43. http://dx.doi.org/10.1016/j.drudis.2011.09.007 PMID: 21945861
[72] Luo H, Li M, Yang M, Wu FX, Li Y, Wang J. Biomedical data and computational models for drug repositioning: A comprehensive review. Brief Bioinform 2021; 22(2): 1604-19. http://dx.doi.org/10.1093/bib/bbz176 PMID: 32043521
[73] Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL. Recent developments of the Chemistry Development Kit (CDK) - an open-source java library for chemo- and bioinformatics. Curr Pharm Des 2006; 12(17): 2111-20. http://dx.doi.org/10.2174/138161206777585274 PMID: 16796559
[74] Hattori M, Okuno Y, Goto S, Kanehisa M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. J Am Chem Soc 2003; 125(39): 11853-65. http://dx.doi.org/10.1021/ja036030u PMID: 14505407
[75] Kashima H, Tsuda K, Inokuchi A. Marginalized kernels between labeled graphs. Proceedings of the 20th international conference on machine learning (ICML-03); 2003 August 21-24; Washington DC, USA.
[76] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform 2015; 7(1): 20. http://dx.doi.org/10.1186/s13321-015-0069-3 PMID: 26052348
[77] Klambauer G, Wischenbart M, Mahr M, Unterthiner T, Mayr A, Hochreiter S. RCHEMCPP: A web service for structural analoging in ChEMBL, Drugbank and the Connectivity Map. Bioinformatics 2015; 31(20): 3392-4. http://dx.doi.org/10.1093/bioinformatics/btv373 PMID: 26088801
[78] Resnik P. Using information content to evaluate semantic similarity in a taxonomy. Proceedings of the 14th International Joint Conference on Artificial Intelligence; 1995 Aug. 20-25; CA, USA.
[79] Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science 2008; 321(5886): 263-6. http://dx.doi.org/10.1126/science.1158140 PMID: 18621671
[80] Takarabe M, Kotera M, Nishimura Y, Goto S, Yamanishi Y. Drug target prediction using adverse event report systems: A pharmacogenomic approach. Bioinformatics 2012; 28(18): i611-8. http://dx.doi.org/10.1093/bioinformatics/bts413 PMID: 22962489
[81] Cheng L, Li J, Ju P, Peng J, Wang Y. SemFunSim: A new method for measuring disease similarity by integrating semantic and gene functional association. PLoS One 2014; 9(6): e99415. http://dx.doi.org/10.1371/journal.pone.0099415 PMID: 24932637
[82] Menche J, Sharma A, Kitsak M, *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 2015; 347(6224): 1257601. http://dx.doi.org/10.1126/science.1257601 PMID: 25700523
[83] Yu G, Wang LG, Yan GR, He QY. DOSE: An R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 2015; 31(4): 608-9. http://dx.doi.org/10.1093/bioinformatics/btu684 PMID: 25677125
[84] Mathur S, Dinakarpandian D. Finding disease similarity based on implicit semantic similarity. J Biomed Inform 2012; 45(2): 363-71. http://dx.doi.org/10.1016/j.jbi.2011.11.017 PMID: 22166490
[85] Paik H, Heo HS, Ban HJ, Cho SB. Unraveling human protein interaction networks underlying co-occurrences of diseases and pathological conditions. J Transl Med 2014; 12(1): 99. http://dx.doi.org/10.1186/1479-5876-12-99 PMID: 24731539
[86] Smith SB, Dampier W, Tozeren A, Brown JR, Magid-Slav M. Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis. PLoS One 2012; 7(3): e33174. http://dx.doi.org/10.1371/journal.pone.0033174 PMID: 22432004
[87] Palme J, Hochreiter S, Bodenhofer U. KeBABS: An R package for kernel-based analysis of biological sequences. Bioinformatics 2015; 31(15): 2574-6. http://dx.doi.org/10.1093/bioinformatics/btv176 PMID: 25812745
[88] Resnik P. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. J Artif Intell Res 1999; 11: 95-130. http://dx.doi.org/10.1613/jair.514

[89] Vapnik V. The nature of statistical learning theory. Springer science & business media 2013.
[90] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 1991; 21(3): 660-74. http://dx.doi.org/10.1109/21.97458
[91] Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemometr 2004; 18(6): 275-85. http://dx.doi.org/10.1002/cem.873
[92] Breiman L. Bagging predictors. Mach Learn 1996; 24(2): 123-40. http://dx.doi.org/10.1007/BF00058655
[93] Breiman L. Random forests. Mach Learn 2001; 45(1): 5-32. http://dx.doi.org/10.1023/A:1010933404324
[94] Qi Y. Random forest for bioinformatics. Ensemble machine learning. Boston, MA: Springer 2012; pp. 307-23. http://dx.doi.org/10.1007/978-1-4419-9326-7_11
[95] Khoshgoftaar TM, Van Hulse J, Napolitano A. Comparing boosting and bagging techniques with noisy and imbalanced data. IEEE Trans Syst Man Cybern Syst 2010; 41(3): 552-68. http://dx.doi.org/10.1109/TSMCA.2010.2084081
[96] Freund Y, Schapire RE. A decision-theoretic generalization of online learning and an application to boosting. J Comput Syst Sci 1997; 55(1): 119-39. http://dx.doi.org/10.1006/jcss.1997.1504
[97] Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal 2002; 38(4): 367-78. http://dx.doi.org/10.1016/S0167-9473(01)00065-2
[98] Peterson LE. K-nearest neighbor. Scholarpedia 2009; 4(2): 1883. http://dx.doi.org/10.4249/scholarpedia.1883
[99] Wang F, Ding Y, Lei X, Liao B, Wu F. Identifying gene signatures for cancer drug repositioning based on sample clustering. IEEE/ACM Trans Comput Biol Bioinform 2020; 1-13.
[100] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 1982; 79(8): 2554-8. http://dx.doi.org/10.1073/pnas.79.8.2554 PMID: 6953413
[101] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012; 25: 1097-105.
[102] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. Machine Learn. 2012; arXiv:1211.5063.
[103] Vincent P, Larochelle H, Lajoie I, *et al.* Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010; 11(12): 3371-408.
[104] Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks. Commun ACM 2020; 63(11): 139-44. http://dx.doi.org/10.1145/3422622
[105] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. Competition and cooperation in neural nets. Berlin, Heidelberg: Springer 1982; pp. 267-85. http://dx.doi.org/10.1007/978-3-642-46466-9_18
[106] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal 2019; 58: 101552. http://dx.doi.org/10.1016/j.media.2019.101552 PMID: 31521965
[107] Agyemang B, Wu WP, Kpiebaareh MY, Nanor E. Drug-target indication prediction by integrating end-to-end learning and fingerprints. Proceedings of the 16th International Computer Conference on Wavelet Active Media Technology and Information Processing; 2019 Dec 14; Chengdu, China. 2019; pp. 266-72.
[108] Huang L, Luo H, Yang M, Wu FX, Wang J. Drug and disease similarity calculation platform for drug repositioning. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2019 Dec. 9-12; San Diego USA, 2021; pp. 124-9. http://dx.doi.org/10.1109/BIBM47256.2019.8983401
[109] Huang L, Luo H, Li S, Wu FX, Wang J. Drug-drug similarity measure and its applications. Brief Bioinform 2021; 22(4): 1-20. PMID: 33152756
[110] Lipscomb CE. Medical subject headings (MeSH). Bull Med Libr Assoc 2000; 88(3): 265-6. PMID: 10928714
[111] van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome. Eur J Hum Genet 2006; 14(5): 535-42. http://dx.doi.org/10.1038/sj.ejhg.5201585 PMID: 16493445
[112] Bullinaria JA, Levy JP. Extracting semantic representations from word co-occurrence statistics: A computational study. Behav Res Methods 2007; 39(3): 510-26. http://dx.doi.org/10.3758/BF03193020 PMID: 17958162
[113] Chen X, Yan CC, Zhang X, *et al.* WBSMDA: Within and between score for MiRNA-disease association prediction. Sci Rep 2016; 6(1): 21106. http://dx.doi.org/10.1038/srep21106 PMID: 26880032
[114] Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: A method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 2011; 7(1): 496. http://dx.doi.org/10.1038/msb.2011.26 PMID: 21654673
[115] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988; 28(1): 31-6. http://dx.doi.org/10.1021/ci00057a005
[116] Zhou X, Menche J, Barabási AL, Sharma A. Human symptoms-disease network. Nat Commun 2014; 5(1): 4212. http://dx.doi.org/10.1038/ncomms5212 PMID: 24967666
[117] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint. Machine Learn 2016; 2016: arXiv:1609.02907. https://arxiv.org/abs/1609.02907
[118] Wang B, Lyu X, Qu J, Sun H, Pan Z, Tang Z. GNDD: A graph neural network-based method for drug-disease association prediction. IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021 Dec. 9-12; San Diego USA, 2019; pp. 1253-5. http://dx.doi.org/10.1109/BIBM47256.2019.8983257
[119] Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug-disease associations through layer attention graph convolutional network. Brief Bioinform 2021; 22(4): 1-11. http://dx.doi.org/10.1093/bib/bbaa243 PMID: 33078832
[120] Lamb J, Crawford ED, Peck D, *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 2006; 313(5795): 1929-35. http://dx.doi.org/10.1126/science.1132939 PMID: 17008526
[121] Lamb J. The connectivity map: A new tool for biomedical research. Nat Rev Cancer 2007; 7(1): 54-60. http://dx.doi.org/10.1038/nrc2044 PMID: 17186018
[122] Wei WQ, Cronin RM, Xu H, Lasko TA, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. J Am Med Inform Assoc 2013; 20(5): 954-61. http://dx.doi.org/10.1136/amiajnl-2012-001431 PMID: 23576672
[123] Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. J Biomed Inform 2017; 70: 54-64. http://dx.doi.org/10.1016/j.jbi.2017.04.021 PMID: 28465082
[124] Yan C, Duan G, Zhang Y, Wu FX, Pan Y, Wang J. Predicting drug-drug interactions based on integrated similarity and semi-supervised learning. IEEE/ACM Trans Comput Biol Bioinform 2020; 2020: 1-12.
[125] Bi X, Ma H, Li J, Ma Y, Chen D. A positive and unlabeled learning framework based on extreme learning machine for drug-drug interactions discovery. J Ambient Intell Humaniz Comput 2018; 22: 1-2. http://dx.doi.org/10.1007/s12652-018-0960-7
[126] Huang G, Song S, Gupta JN, Wu C. Semi-supervised and unsupervised extreme learning machines. IEEE Trans Cybern 2014; 44(12): 2405-17. http://dx.doi.org/10.1109/TCYB.2014.2307349 PMID: 25415946
[127] Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. Bioinformatics 2018; 34(7): 1164-73. http://dx.doi.org/10.1093/bioinformatics/btx731 PMID: 29186331
[128] Wang B, Mezlini AM, Demir F, *et al.* Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 2014; 11(3): 333-7. http://dx.doi.org/10.1038/nmeth.2810 PMID: 24464287

[129] Zhou B, Wang R, Wu P, Kong DX. Drug repurposing based on drug-drug interaction. Chem Biol Drug Des 2015; 85(2): 137-44. http://dx.doi.org/10.1111/cbdd.12378 PMID: 24934184

[130] Munir A, Elahi S, Masood N. Clustering based drug-drug interaction networks for possible repositioning of drugs against EGFR mutations: Clustering based DDI networks for EGFR mutations. Comput Biol Chem 2018; 75: 24-31. http://dx.doi.org/10.1016/j.compbiolchem.2018.04.011 PMID: 29730365

[131] Polikar R. Ensemble based systems in decision making. IEEE Circuits Syst Mag 2006; 6(3): 21-45. http://dx.doi.org/10.1109/MCAS.2006.1688199

[132] Peng B, Ning X. Deep learning for high-order drug-drug interaction prediction. Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. Niagara Falls USA. 2019; pp. 197-206. http://dx.doi.org/10.1145/3307339.3342136

[133] Lin X, Quan Z, Wang ZJ, Ma T, Zeng X. KGNN: Knowledge graph neural network for drug-drug interaction prediction. Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama Japan. 2020; pp. 2739-45. http://dx.doi.org/10.24963/ijcai.2020/380

[134] Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician 2007; 76(3): 391-6. https://www.aafp.org/afp/2007/0801/p391.html PMID: 17708140

[135] Chu Y, Kaushik AC, Wang X, *et al.* DTI-CDF: A cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. Brief Bioinform 2021; 22(1): 451-62. http://dx.doi.org/10.1093/bib/bbz152 PMID: 31885041

[136] Zhou ZH, Feng J. Deep forest Machine Learn 2017; 2017: arXiv: 1702.08835. https://arxiv.org/abs/1702.08835

[137] Lin YT, Sheu SY, Lin CC. Prediction of drug-protein interaction and drug repositioning using machine learning model. bioRxiv 2020; 2020: 218826v1. https: //www.biorxiv.org/content/10.1101/2020.07.29.218826v1

[138] Zong N, Kim H, Ngo V, Harismendy O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. Bioinformatics 2017; 33(15): 2337-44. http://dx.doi.org/10.1093/bioinformatics/btx160 PMID: 28430977

[139] Cheng F, Liu C, Jiang J, *et al.* Prediction of drug-target interactions and drug repositioning *via* network-based inference. PLOS Comput Biol 2012; 8(5): e1002503. http://dx.doi.org/10.1371/journal.pcbi.1002503 PMID: 22589709

[140] Liu H, Sun J, Guan J, Zheng J, Zhou S. Improving compound-protein interaction prediction by building up highly credible negative samples. Bioinformatics 2015; 31(12): i221-9. http://dx.doi.org/10.1093/bioinformatics/btv256 PMID: 26072486

[141] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models *via* coordinate descent. J Stat Softw 2010; 33(1): 1-22. http://dx.doi.org/10.18637/jss.v033.i01 PMID: 20808728

[142] You J, McLeod RD, Hu P. Predicting drug-target interaction network using deep learning model. Comput Biol Chem 2019; 80: 90-101. http://dx.doi.org/10.1016/j.compbiolchem.2019.03.016 PMID: 30939415

[143] Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J Comput Chem 2011; 32(7): 1466-74. http://dx.doi.org/10.1002/jcc.21707 PMID: 21425294

[144] Kawashima S, Kanehisa M. AAindex: Amino acid index database. Nucleic Acids Res 2000; 28(1): 374-4. http://dx.doi.org/10.1093/nar/28.1.374 PMID: 10592278

[145] Michel M, Menéndez Hurtado D, Elofsson A. PconsC4: Fast, accurate and hassle-free contact predictions. Bioinformatics 2019; 35(15): 2677-9. http://dx.doi.org/10.1093/bioinformatics/bty1036 PMID: 30590407

[146] Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation. J Chem Inf Model 2019; 59(9): 3981-8. http://dx.doi.org/10.1021/acs.jcim.9b00387 PMID: 31443612

[147] Subramanian A, Narayan R, Corsello SM, *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017; 171(6): 1437-1452.e17. http://dx.doi.org/10.1016/j.cell.2017.10.049 PMID: 29195078

[148] Szklarczyk D, Gable AL, Lyon D, *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47(D1): D607-13. http://dx.doi.org/10.1093/nar/gky1131 PMID: 30476243

[149] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27(12): 1739-40. http://dx.doi.org/10.1093/bioinformatics/btr260 PMID: 21546393

[150] Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 2020; 18: 784-90. http://dx.doi.org/10.1016/j.csbj.2020.03.025 PMID: 32280433

[151] Wang M, Cao R, Zhang L, *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. Cell Res 2020; 30(3): 269-71. http://dx.doi.org/10.1038/s41422-020-0282-0 PMID: 32020029

[152] Liu B, Dai Y, Li X, Lee WS, Yu PS. Building text classifiers using positive and unlabeled examples. Third IEEE International Conference on Data Mining; 2003 Nov, 22; Melbourne, FL, USA. http://dx.doi.org/10.1109/ICDM.2003.1250918

[153] Lan W, Wang J, Li M, *et al.* Predicting drug–target interaction using positive-unlabeled learning. Neurocomputing 2016; 206: 50-7. http://dx.doi.org/10.1016/j.neucom.2016.03.080

[154] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019; 1(5): 206-15. http://dx.doi.org/10.1038/s42256-019-0048-x