


Drug–drug interaction prediction: databases, web servers and computational models

Yan Zhao, Jun Yin, Li Zhang, Yong Zhang and Xing Chen 

Corresponding author: Xing Chen, School of Science, Jiangnan University, Wuxi 214122, China. E-mail: xingchen@amss.ac.cn; 8202305001@jiangnan.edu.cn

Abstract

In clinical treatment, two or more drugs (i.e. drug combination) are simultaneously or successively used for therapy with the purpose of primarily enhancing the therapeutic efficacy or reducing drug side effects. However, inappropriate drug combination may not only fail to improve efficacy, but even lead to adverse reactions. Therefore, according to the basic principle of improving the efficacy and/or reducing adverse reactions, we should study drug–drug interactions (DDIs) comprehensively and thoroughly so as to reasonably use drug combination. In this review, we first introduced the basic conception and classification of DDIs. Further, some important publicly available databases and web servers about experimentally verified or predicted DDIs were briefly described. As an effective auxiliary tool, computational models for predicting DDIs can not only save the cost of biological experiments, but also provide relevant guidance for combination therapy to some extent. Therefore, we summarized three types of prediction models (including traditional machine learning-based models, deep learning-based models and score function-based models) proposed during recent years and discussed the advantages as well as limitations of them. Besides, we pointed out the problems that need to be solved in the future research of DDIs prediction and provided corresponding suggestions.

Keywords: drug; drug–drug interaction; database; web server; computational model

Drug

In pharmacology, as a class of chemical substance with known structure, drugs can produce biological effect when they are administered to a living organism [1]. More specifically, a pharmaceutical drug, also known as a medication or medicine, is a chemical substance used to prevent or treat diseases [1]. Unlike food, for patients with different diseases, drugs may be taken into the bodies in different ways, such as inhalation, injection, ingestion, skin application, sublingual dissolution and so on [2]. Besides, in clinical treatment, patients usually take drugs for a limited period of time or periodically over a long period of time [3]. With the development of science and technology, the way of manufacturing drugs has changed a lot [4]. Traditionally, drugs are derived from medicinal plants, but they have also been synthesized organically in recent years [5]. As the number of drugs increase, a number of relevant databases have been constructed for further research [6, 7]. For example, the latest release of Drug-Bank [6] (version 5.1.10, released 4 January 2023) contains 15 451 drug entries including 2740 approved small molecule drugs, 1577 approved biologics, 134 nutraceuticals and over 6717 experimental drugs.

Drug–drug interaction

In clinical treatment, to cure or ameliorate symptoms of diseases or medical condition, two or more kinds of drugs are usually needed [8]. The fundamental reason of drug combination lies in the evidence that combination therapy tends to have higher cure rates than monotherapy [9]. For instance, the combination therapy of doxorubicin, cyclophosphamide, vincristine and prednisone is commonly implemented in cancer chemotherapy regimens [10, 11]. Another example is that the combined application of anti-tuberculosis drugs not only enhances the drug efficacy but also delays the emergence of drug resistance of *Mycobacterium tuberculosis*, which is the main cause of tuberculosis [12]. However, the risk of harmful drug–drug interactions (DDIs) increases as patients take more kinds of drugs [13]. For example, more than one-third of older Americans regularly use five or more drugs or supplements, and 15% are at risk for serious DDIs [14]. Specifically, DDIs refer to the interactions between drugs used in drug therapy or between drugs and metabolites, endogenous substances, food and diagnostic agents [15]. DDIs can either enhance the efficacy (synergistic action) or decrease the efficacy (antagonistic action) by resulting in changes in the nature, intensity, duration, side

Yan Zhao is a PhD student of School of Information and Control Engineering, China University of Mining and Technology. His research interests include bioinformatics, complex network algorithm and machine learning.

Jun Yin is a full-time master student of School of Information and Control Engineering, China University of Mining and Technology. His research interests include bioinformatics, complex network algorithm and machine learning.

Li Zhang is a PhD student of School of Information and Control Engineering, China University of Mining and Technology. His research interests include bioinformatics, drug discovery and the graph neural network.

Yong Zhang, PhD, is a professor of School of Information and Control Engineering, China University of Mining and Technology. His research interests include swarm intelligence, pattern recognition and machine learning.

Xing Chen, PhD, is a professor of School of Science, Jiangnan University. His research interests include complex disease-related non-coding RNA biomarker prediction, computational models for drug discovery and early detection of human complex disease based on big data and artificial intelligence algorithms.

Received: July 21, 2023. **Revised:** October 26, 2023. **Accepted:** November 14, 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

effects and toxicity of drugs [16]. In other words, the consequences of DDIs include the desired, the insignificant (vast majority) and the harmful reaction [17]. In general, we are more concerned about harmful DDIs in terms of drug safety. To improve drug safety, it is necessary to fully understand the pharmacological action of each drug before combination therapy so as to achieve the best curative effect and the least adverse drug reactions [18].

Drugs in combination prescriptions would influence each other's effects and change the way they work in the human body [19]. This kind of influence can be divided into pharmaceutical interactions, pharmacodynamics (PD) interactions and pharmacokinetics (PK) interactions [19]. Pharmaceutical interactions refer to the change of drug action caused by chemical reaction due to unreasonable dispensing, that is, interactions *in vitro* before drugs enter the body [20]. An example of pharmaceutical interaction is that tetracycline and calcium salt injection may result in precipitate due to the formation of chelate under neutral or alkaline conditions [21]. PD interaction means that two drugs share the same receptor, and one drug has antagonistic, additive, synergistic or indirect pharmacological effects on the other drug [13]. For instance, because both atropine and tubocurarine reversibly bind to receptors, the combination therapy of these two drugs blocks the action of the normal physiological transmitter acetylcholine [22]. PK interactions refer to the interference caused by simultaneous or sequential use of two or more drugs in the metabolic stage (absorption, distribution, metabolism and elimination), resulting in enhanced efficacy or adverse reactions [23]. Besides, different from PD interactions, PK interactions often lead to changes in the blood concentration of the interacting drugs [23]. For example, the combination therapy of warfarin and nonsteroidal anti-inflammatory drugs would result in a PK interaction [24]. In detail, some nonsteroidal anti-inflammatory drugs can inhibit warfarin metabolism, thereby enhancing the effect of warfarin on hypoprothrombinemia and significantly increasing the risk of bleeding [24]. For those patients who are receiving combination therapy, unidentified DDIs may reduce efficacy, cause unexpected side effects or other adverse drug reactions and even endanger life [25]. The harmful reactions from DDIs include bleeding, bone marrow suppression, arrhythmias, hypotension, rhabdomyolysis, central nervous system depression, seizures, hypoglycemia, renal failure and so on [26].

Hence, paying attention to DDIs is of great significance to adopt effective combination therapy and further improve the quality of medical treatment. Moreover, in-depth understanding of the absorption, distribution, metabolism and excretion process of drugs in the body, as well as the interactions between various drugs in the body, can reduce adverse drug reactions and ensure drug safety [27]. Nevertheless, as a new drug is approved, the potential DDIs resulting from different drug combinations will grow exponentially. In this case, it is not hard to imagine that the process of validating potential DDIs one by one through biological experiments is very expensive and time-consuming [28]. With the rapid development of computing technology, large-scale DDIs analysis and prediction have also made great strides. In the face of the huge number of existing available drugs, computational models can be used to screen drug combinations with high probability of interaction [16].

As increasing evidence show that the discovery of potential DDIs plays an important role in drug development and disease treatment, more and more researchers devote themselves to corresponding studies. In the following parts, we first reviewed the existing databases and web servers about DDIs. Then, we introduced three types of computational models and discussed their

advantages and disadvantages. Furthermore, we put forward the problems that need to be solved in the future research of DDIs prediction and provided corresponding suggestions.

Databases and web servers

With the rapid development of drug-related research studies, more and more drug-related databases and web servers have been constructed to facilitate researchers to carry out deeper research studies. Next, we briefly introduced some representative databases as well as web servers and summarized them in Table 1.

Databases

We here briefly introduced several well-known databases that were built to store various data related to DDIs such as DrugBank [6], DDInter [29], SuperDRUG2 [30], INXBASE [31], OncoRx [32], DIBD [33], DrugComb [7] and DailyMed. Except for DDIs, users can also retrieve many other useful information about drugs from these databases, including drug target, metabolic pathways, crystal structures, regulatory details, indications, side-effects, physicochemical properties, pharmacokinetics and so on.

DrugBank (<https://www.drugbank.ca/>)

The latest release of DrugBank (version 5.1.10, released on 4 January 2023) contains 15 451 drugs including 2740 approved small molecules, 1577 approved biologics (proteins, peptides, vaccines and allergenics), 134 nutraceuticals and over 6717 experimental drugs [6]. Each drug in DrugBank contains more than 200 data fields with half of the information being devoted to introducing drugs from the aspects of chemistry, pharmacology as well as pharmacy, and the other half being devoted to recording the sequence, structure and pathway of the drug target. In addition, DrugBank records more than 1.3 million DDIs and provides DDIs Checker, through which users can check the interactions between up to five drugs at one time.

DDInter (<http://ddinter.scbdd.com>)

As a comprehensive database dedicated to DDI research, the DDInter not only records 236 834 DDIs involving 1833 drugs but also documents the detailed information about each DDI, such as mechanisms, risk levels, recommendations for drug adjustment and so on [29]. Besides, similar to DrugBank, DDInter also provides Interaction Checker for users to check whether drugs interact with each other.

SuperDRUG2 (<http://cheminfo.charite.de/superdrug2>)

The SuperDRUG2 database is intended to serve as a comprehensive knowledge base of approved and marketed 4587 drugs (involving small molecule, biological products and other drugs) [30]. The annotation of drugs contains regulatory details, chemical structures (2D and 3D), dosage, biological targets, physicochemical properties, external identifiers, side-effects, pharmacokinetic data and DDIs. Besides, SuperDRUG2 can be used to infer potential DDIs and further provide alternative recommendations for elderly patients.

INXBASE (<https://www.medbase.fi/en/professionals/inxbase/>)

INXBASE (formerly named SFINX), a database that records more than 20 000 DDIs, can be easily integrated to health information systems and accessible through a portal, which helps the healthcare professionals choose the most appropriate action to overcome specific DDIs [31]. The database has become a basic

Table 1: The function and URL of databases as well as web servers

Databases or web servers	Function	URL
DrugBank	Recording 15 451 drugs and providing more than 200 data fields for each drug, with half of the information devoted to chemical, pharmacological, pharmaceutical and other aspects of the drug and the other half dedicated to documenting the sequence, structure and pathway of the drug target.	https://www.drugbank.ca/
DDInter	Recording 236 834 DDIs involving 1833 drugs and documenting the detailed information about each DDI, such as mechanisms, risk levels, recommendations for drug adjustment and so on.	http://ddinter.scbdd.com
SuperDRUG2	Recording the annotation of drugs, including regulatory details, chemical structures (2D and 3D), dosage, biological targets, physicochemical properties, external identifiers, side-effects, pharmacokinetic data and DDIs.	http://cheminfo.charite.de/superdrug2
INXBASE	Recording more than 20 000 DDIs	https://www.medbase.fi/en/professionals/inxbase/
OncoRx	Documenting 943 DDIs between 117 ACDs and 166 CAMs.	http://www.onco-informatics.com/oncorx/index.php
DIDB	Recording the drug interaction results derived from drug–drug, drug–food and drug–herb interaction studies.	https://www.druginteractionsolutions.org/
DrugComb	Documenting the standardized results of drug combination screening studies about 739 964 combinations involving 8397 drugs.	https://drugcomb.fimm.fi/
DailyMed	Recording the essential scientific information for the safe and effective use of the drugs, such as indications, dosage, administration, adverse reactions, DDIs and so on.	https://dailymed.nlm.nih.gov/dailymed/about-dailymed.cfm
PolySearch2	Predicting the relationships between biomedical entities, such as human diseases, genes, SNPs, proteins, drugs, metabolites and so on.	http://polysearch.ca
DDI-CPI	Presenting the predicted probabilities of interactions between the given drug and 2515 drugs in the library of DDI-CPI.	http://cpi.bio-x.cn/ddi/
vNN-ADMET	Predicting the ADMET properties of drugs.	https://vnnadmet.bhsai.org/vnnadmet/login.xhtml

tool to avoid DDIs for physicians, pharmacists or nurses. Also, a patient-oriented version is available.

OncoRx (<http://www.onco-informatics.com/oncorx/index.php>)

OncoRx is an oncology database that documents 943 DDIs between 117 anticancer drugs (ACDs) and 166 complementary and alternative medicines (CAMs) [32]. What needs to be pointed out is that OncoRx primarily covers PK and PD DDIs, as these two kinds of interactions explain most of the clinically relevant interactions between drugs. Besides, when users applied OncoRx to search DDIs, some important information would also be provided, such as DDIs parameters, pharmacokinetic data on ACDs and CAMs as well as characteristics of CAMs based on traditional Chinese medicines principles.

DIDB (<https://www.druginteractionsolutions.org/>)

The Drug Interaction Database (DIDB) is designed to support the decision-making process of scientists in evaluating PK DDIs and drug safety, which is composed of human *in vitro* and *in vivo* datasets [33]. The human *in vitro* datasets contain results from both metabolism and transporter studies, while the human *in vivo* datasets include the studies result about organ impairment, pharmacogenetics and drug interaction, where the drug interaction results are derived from drug–drug, drug–food and drug–herb interaction studies.

DrugComb (<https://drugcomb.fimm.fi/>)

As an open-access data portal, DrugComb documents the standardized results of drug combination screening studies about

739 964 combinations involving 8397 drugs [7]. In addition, DrugComb provides a web server, through which users can analyze and visualize their own drug combination screening data.

DailyMed (<https://dailymed.nlm.nih.gov/dailymed/about-dailymed.cfm>)

The DailyMed database contains labels for two types of drugs, namely, FDA-approved drugs (such as prescription drug, nonprescription drug, certain medical devices, etc.) and additional drugs regulated but not approved by the FDA (such as dietary supplements and unapproved prescription as well as nonprescription, etc.). It should be pointed out that the labels of prescription drug and biological products contain a summary of the essential scientific information for the safe and effective use of the product, such as indications, dosage, administration, adverse reactions, DDIs and so on.

Web servers

Except for databases, there are also some online web servers that can be used to analyze or predict DDIs, such as PolySearch2 [34], DDI-CPI [35] and vNN-ADMET [36].

PolySearch2 (<http://polysearch.ca>)

PolySearch2, an online text-mining system, can provide relationships between biomedical entities, such as human diseases, genes, single-nucleotide polymorphisms (SNPs), proteins, drugs, metabolites and so on [34]. Specifically, for one given entity that support retrieval, PolySearch2 will return all types of aforementioned entities associated with this entity. In the search results, each type of entity is sorted in reverse order based on the Z-score calculated by PolySearch [37]. Synonyms of each entity as well as key sentences mined from literatures to confirm the corresponding association

Table 2: The significance and related link of computational models

Model	Significance	Link to the GitHub or sites
Bayesian probabilistic method-based model [13] INDI [57]	Introducing the system connection score and drug phenotypic similarity score Applying a novel scoring scheme to construct the feature vectors of drug pairs based on multiple types of drug similarity	http://www.picb.ac.cn/hanlab/DDI
Label propagation-based model [65] Collective PSL-based model [71] Random forest-based model [73] Logistic regression-based model [81]	Implementing label propagation based on multiple similarity information Applying the hinge-loss MRFs to identify potential DDIs in the multigraph through maximum a posteriori Introducing the enrichment score of the targets of drugs Implementing prediction based on two interaction networks constructed based on the information about PK and PD interactions	
PUL-based model [84] Meta-learning-based model [88] MRMF [93] DDINMF [96]	Applying the growing self-organizing maps clustering algorithm to identify reliable negative samples Using node2vec to get the feature vectors of drugs from the feature network Introducing manifold regularization into matrix factorization Introducing the feature matrix of drug into matrix factorization to make the model suitable for predicting enhancive and depressive DDIs between known drugs and new drugs	
TMFUF [99] LCM-DS [100]	Being suitable for predicting not only known but also new drugs that interact with new drugs Introducing the Dempster-Shafer theory of evidence to integrate the results of three local classification models	https://github.com/justinshi2016/ScientificReports2018
DDIGP [106] Gradient boosting-based model [107] Network algorithm and matrix perturbation algorithm-based model [114] HNAL [118] IAC [121]	Applying the KNNs to fill in the adjacency matrix Using the TPE approach to optimize the hyperparameters of the classifier Applying the classifier ensemble rule to take the logistic regression to map the outputs of all models to a score as the final prediction result Applying five prediction models to identify potential DDIs, respectively Introducing the action crossing method to obtain the feature vectors of drug pairs according to the information about drug-enzyme and drug-transporter actions Introducing the sparse feature learning ensemble method to project drugs from different feature spaces to the common interaction space	https://github.com/zw9977129/drug-drug-interaction/
SFLIN [122]	Applying the DNN to calculate the interaction probabilities based on the feature vectors of drugs	https://github.com/BioMedicalBigDataMiningLabWuhu/SFLIN
DDIMDL [128] SSI-DDI [129] STNN-DDI [132] META-DDIE [135]	Applying the GAT layers to extract the feature vectors of atoms contained in drugs Introducing tensor to describe the interactions between substructures of drugs Introducing the chemical sequential pattern mining algorithm to obtain a set of discrete frequent substructures of drugs	https://github.com/YifanDengWUH/DDIMDL https://github.com/kanz76/SSI-DDI https://github.com/zsy-9/STNN-DDI https://github.com/YifanDengWUH/META-DDIE
DANN-DDI [139] MRCGNN [143] MCFE-MTDDI [146] DSL-DDI [149]	Introducing the structural deep network embedding method to learn the embeddings of drugs from interaction networks Introducing the contrastive learning to obtain the representations of drugs Introducing the extra label-based feature vector to make the model suitable for multi-label prediction Introducing the GNN to extract the substructure representations of drugs	https://github.com/haodandandan/DANN-DDI https://github.com/Zhankun-Xiong/MRCGNN https://github.com/ChendiHan111/MCFE-MTDDI
DSN-DDI [151]	Applying the intra-view and inter-view representation learning methods to obtain the representations of drugs	https://github.com/microsoft/Drug-Interaction-Research/tree/DSN-DDI-for-DDI-Prediction
BioDKG-DDI [154] MDF-SA-DDI [159] Deep feed-forward network-based model [161] R ² -DDI [164] Graph kernel-based approach [167] Semantic prediction-based model [174] Att-BLSTM [177] PM-BLSTM [179] A two-stage DDIs extraction model [181] IK-DDI [186] 3DGT-DDI [189] Russell-Rao-based model [192] Score matrix and PCA-based model [194]	Applying a novel similarity fusion method to fuse multiple similarity matrixes of drugs Introducing the multi-head self-attention mechanism to integrate the feature vectors of each drug pair Introducing the GO term-based drug similarity Applying the MLP to obtain the refinement vectors of drugs Constructing all-path graph kernels to describe the connections between syntactic and semantic within the sentences Introducing four types of semantic prediction generated by SemRep Combining attention mechanism and the RNN with BLSTM to learn the global semantic representation of the sentence Applying a rule to filter the drugs to ensure that only one drug pair in each sentence was studied Applying the SVM classifier to identify DDIs and the LSTM-based classifier to identify the type of DDIs Introducing key external text derived from the DrugBank Introducing the 3D structure conformations of drugs Applying the Russell-Rao method to calculate interaction probability Applying PCA method to integrate the score matrixes to obtain the interaction probability matrix	https://github.com/ShenggenLin/MDF-SA-DDI https://github.com/injic16/R2-DDI https://sbmi.uth.edu/cbb/resources/ddi.htm https://github.com/DouMingLiang/IK-DDI https://github.com/hehh77/3DGT-DDI

are also provided. It should be pointed out that, to improve the accuracy and coverage, the retrieval results presented to users are mined from well-known free-text collections (e.g. MEDLINE, PubMed and Wikipedia) and biological databases (e.g. UniProt and DrugBank).

DDI-CPI (<http://cpi.bio-x.cn/ddi/>)

Considering that a large amount of DDIs are mediated by drug–protein interactions, the DDI-CPI server [35] is constructed to predict potential DDIs based on the chemical–protein interactome (CPI), which is a methodology that mimics the theoretical interactions between drug and proteins using silicon simulations [38]. For a given drug, DDI-CPI will present the predicted probabilities of interactions between the drug and 2515 drugs in the library of DDI-CPI.

vNN-ADMET (<https://vnnadmet.bhsai.org/vnnadmet/login.xhtml>)

Through the vNN-ADMET webserver [36], users can obtain the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of drugs by using one of the fifteen models constructed based on the variable nearest neighbor (vNN) method [39]. For example, these models could be applied to predict some important properties of the given drug, such as cytotoxicity, mutagenicity, cardiotoxicity, DDIs, microsomal stability and drug-induced liver injury.

Computational models

As mentioned above, detecting DDIs is beneficial to clinical drug combination treatment. However, due to the high cost and long cycle of experimental methods, it is of great significance to develop effective computational models to infer potential DDIs on a large scale. During recent years, to predict unknown DDIs, researchers have built a number of computational models, which could be divided into three categories: traditional machine learning-based models, deep learning-based models and score function-based models.

Traditional machine learning-based models

Traditional machine learning algorithms have been widely used to solve complex problems in industrial application [40, 41] and biological science [42–52]. Here, traditional machine learning-based prediction models mainly covers label propagation, Markov random fields (MRFs), random forest, logistic regression, support vector machine (SVM), matrix factorization, ensemble learning and so on. Traditional machine learning-based models could be used to predict DDIs on a large scale and are suitable for new drugs. However, there are still some limitations to be resolved. For example, in the models constructed based on supervised learning algorithms, unlabeled samples are treated as negative samples because of the lack of highly reliable negative samples. Besides, for the parameters involved in the traditional machine learning-based models, researchers often randomly set the values of the parameters rather than using some algorithms to obtain the optimal values of the parameters, which limits the performance of models to some extent. Moreover, researchers tended to obtain the feature vectors of drug pairs by splice the feature vectors of corresponding drugs directly, so constructing more significant feature vectors is still an urgent problem to be solved.

Bayesian probabilistic method-based model

Based on the hypothesis that the smaller the minimum distance between the targets of two drugs in the PPI network constructed

based on the Human Protein Reference Database [53], the greater the possibility of PD interaction between the corresponding drugs, Huang et al. [13] designed a model for PD interaction prediction by considering drug actions in the PPI network (Figure 1). Firstly, for protein p_i , the authors constructed its coding gene's expression profile across 79 human tissues [54], denoted by EP_i (79D vector). Secondly, for the protein pair (p_i, p_j) with known interaction, the authors calculated the Pearson correlation coefficient between EP_i and EP_j to weight the edge connecting p_i and p_j in the PPI network. Besides, for each drug, the authors constructed a target-centered system consisting of the target proteins of the corresponding drug and the first-step neighboring proteins of the target protein in the PPI network. Finally, for drugs d_i and d_j , the system connection score $S\text{-score}_{ij}$ was calculated to describe the tightness of connection between the target-centered systems of d_i and d_j :

$$S\text{-score}_{ij} = \frac{\bar{x}_{ij} - \mu_0}{s_{ij}/\sqrt{n_{ij}}} \quad (1)$$

where \bar{x}_{ij} and s_{ij} represent the mean and SD of edge weights connecting the proteins in the target-centered system of d_i and d_j , respectively; n_{ij} denotes the number of edges connecting two target-centered systems; and μ_0 refers to the mean of all edge weights in the PPI network. It should be noted that if two target-centered systems had a common protein, an artificial edge with a weight of 1 was added between the two systems. In addition, following the previous research [55], the drug phenotypic similarity score $P\text{-score}_{ij}$ between d_i and d_j was calculated based on the clinical side effects of drugs. Finally, inspired by the Bayesian probabilistic model proposed by Xia et al. [56], the likelihood ratio (LR) for drug pair (d_i, d_j) to be true-positive DDIs versus true-negative DDIs based on $S\text{-score}_{ij}$ and $P\text{-score}_{ij}$ were calculated:

$$LR(S\text{-score}_{ij}) = \frac{P(S\text{-score}_{ij}|\text{positive})}{P(S\text{-score}_{ij}|\text{negative})} \quad (2)$$

$$LR(P\text{-score}_{ij}) = \frac{P(P\text{-score}_{ij}|\text{positive})}{P(P\text{-score}_{ij}|\text{negative})} \quad (3)$$

Then, by multiplying the LR calculated based on two independent evidences (i.e. $S\text{-score}_{ij}$ and $P\text{-score}_{ij}$), the interaction probability P_{ij} between d_i and d_j was obtained:

$$P_{ij} = LR(S\text{-score}_{ij}) \times LR(P\text{-score}_{ij}) \quad (4)$$

INDI

Instead of predicting a single type of DDIs, the model of INferring Drug Interactions (INDI) proposed by Gottlieb et al. [57] could be used to infer PD, PK and potential PK interactions (the drug pair was metabolized by the same cytochrome P450 enzyme but there was no interaction evidence between these two drugs). Firstly, the authors collected these three types of DDIs from DrugBank [6] and Drugs.com (<http://drugs.com>). Then, they calculated drug similarity from seven aspects, including the chemical structure [58], receptors [59], side effects [60] and anatomical therapeutic chemical (ATC) codes [61] of drugs, the sequence of drug targets [62], the distances between drug targets on the human PPI network [63] and the semantic similarity between drug targets [61], which were denoted by S_i ($i=1,2,\dots,6,7$), respectively. Further, they integrated the above multiple drug–drug similarities to obtain the features of drug pairs. Specifically, given a drug pair (d_1, d_2) without

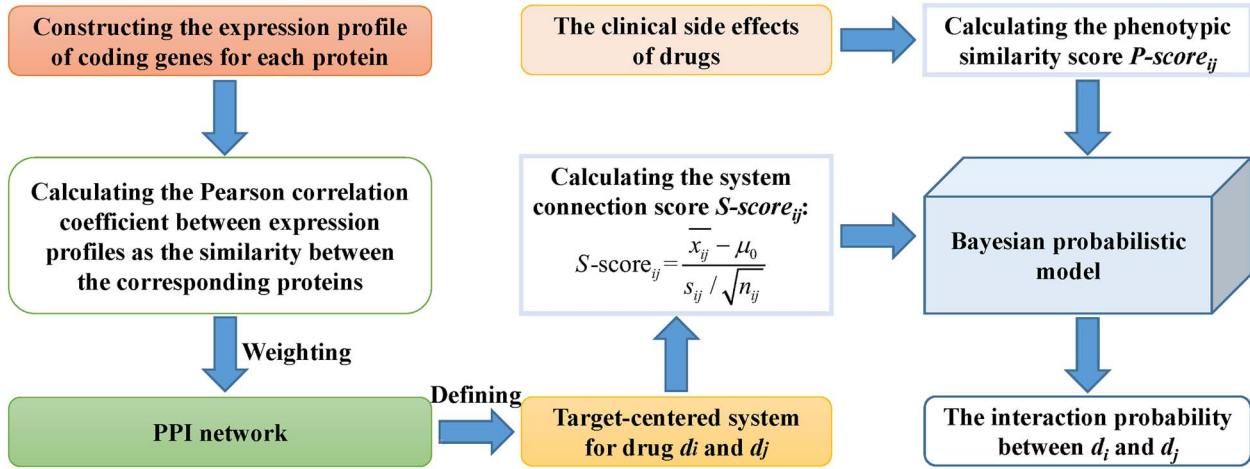


Figure 1. The flowchart of Bayesian probabilistic method-based model, where the interaction probabilities between drugs are calculated by integrating the system connection score and phenotypic similarity score through a Bayesian probabilistic method.

verified interaction, for each known interaction (d'_1, d'_2) , they first computed the drug-drug similarities $S_i(d_1, d'_1), S_i(d_2, d'_2), S_i(d_1, d'_2)$ and $S_i(d_2, d'_1)$. Next, according to the scoring scheme proposed in literature [64], they calculated the score $Score(d_1, d_2)$ between d_1 and d_2 as follows:

$$S_{ij}(d_1, d_2, d'_1, d'_2) = \max \left(\sqrt{S_i(d_1, d'_1) \times S_j(d_2, d'_2)}, \sqrt{S_i(d_1, d'_2) \times S_j(d_2, d'_1)} \right) \quad (5)$$

$$Score_{ij}(d_1, d_2) = \max_{\substack{(d'_1, d'_2) \cap \{d_1, d_2\} = \emptyset \\ (d'_1, d'_2) \in D_p}} S_{ij}(d_1, d_2, d'_1, d'_2) \quad (6)$$

where D_p represents the set consisting of drug pairs with known interactions; $i, j = 1, 2, \dots, 6, 7$. Therefore, 49 scores [i.e. $Score_{11}(d_1, d_2), Score_{12}(d_1, d_2), \dots, Score_{76}(d_1, d_2), Score_{77}(d_1, d_2)$] between d_1 and d_2 could be obtained, which are regarded as the features of the drug pair (d_1, d_2) . Finally, a logistic regression classifier was implemented to predict three types of DDIs based on the features of drug pairs.

Label propagation-based model

By integrating the information about side effects and chemical structure of drugs, Zhang et al. [65] proposed a model to predict potential DDIs based on the label propagation algorithm. Firstly, the authors downloaded the information about the label and off-label side effect of drugs from SIDER [66] and OFFSIDES [67], respectively. Besides, the chemical structure information was extracted from PubChem [68]. Secondly, for each drug, three binary feature vectors were constructed according to the information about label side effect, off-label side effect and chemical structure, respectively. Then, the Jaccard index was used to calculate the similarities between drugs and three corresponding similarity matrices were constructed later. Thirdly, the authors used the Bregmanian Bi-Stochasticity (BBS) algorithm [69] to normalize the three similarity matrices and the normalized matrixes were denoted as $W_k (k = 1-3)$. Besides, the adjacency matrix A was constructed based on the known DDIs, whose row was defined as the label of the corresponding drug. As for label propagation, in the t th iteration, the drug nodes absorbed the label information of the neighbor nodes at a ratio of μ and retained the original label information with a ratio of $(1 - \mu)$ to update the label. Therefore,

based on the k th similarity matrix W_k , the interaction probability matrix P_k^t could be obtained as follows:

$$P_k^t = \mu W_k P_k^{t-1} + (1 - \mu) A \quad (7)$$

where $P_k^0 = A$. The matrix P_k^t obtained after iteration convergence is the final interaction probability matrix, which could also be obtained by minimizing the following objective function:

$$\min_{P_k} \mu \text{tr}(P_k^T (I - W_k) P_k) + (1 - \mu) \|P_k - A\|_F^2 \quad (8)$$

where $\text{tr}()$ refers to the trace of a matrix; $\|\cdot\|_F$ refers to the Frobenius norm. To introduce multiple similarity information about drugs, the authors calculated the converged solution by solving the following composite optimization problem:

$$\begin{aligned} \min_{P, \alpha} \mu \text{tr} \left(P^T \sum_{k=1}^3 \alpha_k (I - W_k) P \right) + (1 - \mu) \|P - A\|_F^2 + \delta \|\alpha\|_2^2 \\ \text{s.t., } \forall k, \alpha_k \geq 0, \sum_{k=1}^3 \alpha_k = 1 \end{aligned} \quad (9)$$

where δ refers to the regularization parameter; $\alpha = [\alpha_1, \alpha_2, \alpha_3]$ represents the vector composed of weight coefficients; $\|\cdot\|_2$ represents the Euclidean norm. Finally, the block coordinate descent (BCD) [70] schema was applied to calculate P and α , where the element of matrix P was regarded as the interaction probability between the corresponding two drugs.

Collective probabilistic soft logic-based model

Through the collective probabilistic soft logic (PSL) framework, Sridhar et al. [71] inferred potential DDIs based on multiple drug similarities and known DDIs. Specifically, the authors calculated seven kinds of drug similarity, including chemical structure-based [58], ligand-based [59], side effect-based [60], drug annotation-based [61], target protein sequence-based [62], PPI network-based [63] and Gene Ontology (GO)-based [61] drug similarities. Besides, by representing drugs with nodes, the authors constructed a multigraph with eight types of edges (denoting the above seven types of drug similarity and known DDIs, respectively). Then, first order logic-syntax was used by PSL to template for a special class of MRFs models called hinge-loss MRFs (HL-MRFs). Finally, HL-MRFs was applied to identify potential DDIs in the multigraph

through maximum a posteriori (MAP) [72] based on the rule that if drug d_i and d_j were similar, and there was known DDI between d_j and d_k , d_i was likely to interact with d_k .

Random forest-based model

Liu *et al.* [73] proposed a random forest-based method to predict unknown DDIs. Before training the random forest model, each drug pair was represented by a feature vector derived from three aspects: chemical interaction between drugs [74], protein interactions between the targets of drugs and enrichment score of the targets of drugs. Taking the drug pair (d_i, d_j) as an example, the feature based on the chemical interaction referred to the ‘Combined_score’ of the drug pair recorded in STITCH [75]. According to the protein interaction score documented in STRING [76], the features of the drug pairs were defined based on different target protein sets and the same target protein set, respectively. Specifically, the author took the maximum DS_{ij}^m and average value DS_{ij}^a of the interaction scores between the target proteins of d_i and the target proteins of d_j as the features built based on a different target protein set. As for the features based on the same target protein set, for these proteins in the target protein set of drug d_i , the authors calculated the maximum SS_i^m and average value SS_i^a of the interaction scores between these proteins. In the same way, the maximum SS_j^m and average value SS_j^a of the interaction scores between the target proteins of the drug d_j were obtained. Then, four features of the drug pair (d_i, d_j) were defined (i.e. $SS_i^m + SS_j^m, SS_i^a + SS_j^a, |SS_i^m - SS_j^m|, |SS_i^a - SS_j^a|$). The authors also constructed the feature vectors of drug pairs based on the enrichment score of target proteins in 229 pathways in Kyoto Encyclopedia of Genes and Genomes (KEGG) [77]. Specifically, for drugs d_i and d_j , the authors calculated their target enrichment scores in these pathways, respectively, denoted by $es_i^1, es_i^2, \dots, es_i^{229}, es_j^1, es_j^2, \dots, es_j^{229}$. Then, according to the enrichment scores, 458 features of drug pair (d_i, d_j) were defined (i.e. $es_i^1 + es_j^1, es_i^2 + es_j^2, \dots, es_i^{229} + es_j^{229}, |es_i^1 - es_j^1|, |es_i^2 - es_j^2|, \dots, |es_i^{229} - es_j^{229}|$). Minimum redundancy maximum relevance [78] as well as incremental feature selection [78] were used to implement feature extraction, and the final 386 features were selected for the drug pair based on the value of the Matthews’s correlation coefficient [79]. Finally, the random forest algorithm with its default configuration (in Weka 3.6.4 [80]) was adopted to train the prediction model.

Logistic regression-based model

Based on the assumption that a query drug (Dq) tends to interact with a drug to be examined (De) if Dq is structurally similar to drugs in the interaction network of De, Takeda *et al.* [81] proposed a DDI prediction model highly relying on the 2D structural similarities between Dq and all drugs in the interaction network of De. Firstly, for each De, the authors constructed two interaction networks based on the PK and PD information about it. The nodes in the network constructed based on PK information represented enzymes as well as transfer proteins associated with De and drugs associated with the above-mentioned enzymes as well as transfer proteins. It should be pointed out that the drugs in the network could be divided into three categories, that is, drugs that interact with enzymes, drugs that have pharmacogenetic associations with enzymes and drugs that are transported by transfer proteins. While the nodes in the network constructed based on PD information refer to the target protein of De, the drugs associated with the target protein, other proteins that interact with the target protein and the drugs associated with these proteins. Similarly,

the drugs in the network are also divided into three categories, namely, drugs that target the target protein, drugs that have pharmacogenetic association with the target protein and drugs having pharmacogenetic association with the protein interacted with target proteins. Secondly, according to the PubChem 2D fingerprint [82] and the Tanimoto coefficient, the structural similarities between Dq and all the drugs (including De) in the two networks of De were computed. Then, they calculated the maximum similarity between Dq and each type of drugs, respectively, which constituted the feature vector (seven-dimensions) of drug pair (Dq, De) together with the similarity between Dq and De. After constructing the balanced classification dataset [the number of (Dq, De) pairs with known interactions is the same as the number of pairs without interactions], the authors trained the logistic regression model by using the generalized linear models (glm) implemented in R package *caret* based on the feature vectors of drug pairs [83]. Finally, they applied the trained logistic regression model to predict the potential DDIs.

Positive-unlabeled learning-based model

To deal with the problem of rarely available negative samples, Hameed *et al.* [84] proposed a positive-unlabeled learning (PUL) method to infer potential DDIs. Initially, based on the chemical structure [6], indication [85], target protein [6] and side effect [86] information of drugs, four binary vectors f_k ($k = 1, 2, 3, 4$) were constructed as the feature vectors of the corresponding drug, respectively. Then, based on these feature vectors, two types of feature vectors of drug pairs were defined. Specifically, the first type was Jaccard index-based similarity feature representation 1 (SFR1), where the drug similarity matrix S_k ($k = 1, 2, 3, 4$) was calculated based on the k th feature vector using the Jaccard index. Therefore, according to SFR1, the author could obtain the 4D feature vector F_1^{ij} of drug pair (d_i, d_j) :

$$F_1^{ij} = (S_1(d_i, d_j), S_2(d_i, d_j), S_3(d_i, d_j), S_4(d_i, d_j)) \quad (10)$$

Besides, the authors raised the similarity feature representation 2 (SFR2) to capture the shared properties of drugs. Specifically, the author averaged the corresponding element values of the k th feature vectors of d_i and d_j , respectively, to obtain the vector F_{2k}^{ij} . Then, the authors defined the feature vector F_2^{ij} constructed based on SFR2:

$$F_2^{ij} = (F_{21}^{ij}, F_{22}^{ij}, F_{23}^{ij}, F_{24}^{ij}) \quad (11)$$

Secondly, the authors considered 6036 drug pairs with interactions recorded in DrugBank as positive samples, and another 6036 samples were randomly selected from unlabeled samples as candidate samples. Then, they utilized the growing self-organizing maps (GSOM) clustering algorithm [87] to cluster the above samples based on SFR1. If a cluster contained only candidate samples, these candidate samples were regarded as negative samples. In a similar way, negative samples were inferred based on SFR2, and the 589 common negative samples inferred based on SFR1 and SFR2 were considered as the final negative samples. Thirdly, they randomly selected 589 samples from the positive samples as the final positive samples, which were combined with the inferred negative samples to form the training set. The authors repeated the sampling 10 times to construct 10 balanced training sets. Then, the authors applied these training sets to train SVM classifiers based on SFR1 and SFR2, respectively. Finally, the prediction

results of the trained 20 classifiers were averaged to obtain the final prediction result.

Meta-learning-based model

Similar to the above PUL-based approach, considering that it is difficult to obtain reliable negative samples, Deepika et al. [88] proposed a semi-supervised learning framework (Figure 2) for predicting DDIs through combining representation learning [89], PUL [90] and meta-learning [91]. Specifically, the authors first applied the same four types of drug features as the above PUL-based approach to construct corresponding feature network, respectively. Taking the feature network constructed based on the chemical structure as an example, there were two kinds of nodes in the network: drug nodes and chemical substructure nodes. If the drug has a certain substructure, the corresponding two nodes are connected by an edge; otherwise, they are not connected, and the other three feature networks were constructed in a similar way. For each type of feature, drugs in the feature network were represented with a d -dimensional feature vector via node2vec [92], which was a representation learning algorithm that was applied to explore the neighborhood information of nodes in the network through biased random walk and then obtain the features of the nodes. For the drug pair (d_1, d_2) , a d -dimensional vector composed of the absolute value of the difference between the corresponding element in the feature vector of d_1 and d_2 was defined as the feature vector of the drug pair. Next, the authors trained the base classifiers based on each type of feature vector and calculated the weight of each base classifier by cross validation, respectively. Then, the multiplication of the interaction probability predicted by the base classifier and the weight of the classifier was defined as the score of corresponding drug pair. Therefore, based on the four types of features, four scores could be obtained. Then, the final feature vectors (four-dimension) of drug pairs were constructed based on their scores to train the meta-classifier to predict drug pairs with potential interaction from unlabeled samples. It should be pointed out that the bagging SVM classifier constructed based on PUL method was used as the base classifier and meta-classifier of this model.

Manifold regularized matrix factorization

Zhang et al. [93] presented a novel computational method named manifold regularized matrix factorization (MRMF) to predict potential DDIs by introducing the drug feature-based manifold regularization into the matrix factorization. The authors defined the adjacency matrix A based on the known DDIs, where if the drug d_i and d_j have known interaction, then $A_{ij} = 1$, otherwise $A_{ij} = 0$. Then, eight feature vectors of drugs were constructed based on the information about the substructures, targets, enzymes, transporters, pathways, indications, side effects and off side effects of drugs, respectively. Further, the Jaccard similarity matrix (S_{jar}), cosine similarity matrix (S_{cos}) and Gauss similarity matrix (S_{gau}) of drugs were calculated based on the feature vectors, respectively. For matrix factorization, in order to approximate the matrix A , two low-rank matrices X and Y could be obtained by minimizing the following objective function:

$$L = \frac{1}{2} \|A - XY^T\|_F^2 + \frac{\lambda}{2} (\|X\|_F^2 + \|Y\|_F^2) \\ = \frac{1}{2} \sum_{ij} (A_{ij} - x_i y_j^T)^2 + \frac{\lambda}{2} \left(\sum_i \|x_i\|_2^2 + \sum_j \|y_j\|_2^2 \right) \quad (12)$$

where $\|A - XY^T\|_F^2$ represents the least square cost function, which is used to ensure that the final product of X and Y

approximates the matrix A ; $(\|X\|_F^2 + \|Y\|_F^2)$ is used to overcome the overfitting problem; x_i refers to the i th row of the matrix X and y_j is the j th row of the matrix Y ; λ represents the Tikhonov regularization parameter. Given the wide applications of manifold learning [94, 95], the authors treated the similarity between drugs as manifolds and assumed that drugs approximately maintained manifolds in the low-dimensional space. Then, the manifold regularizations for drugs in the low-dimensional space were defined as follows:

$$L_{reg}^{row} = \frac{1}{2} \sum_{ij} \|x_i - x_j\|_2^2 S_{ij} \quad (13)$$

$$L_{reg}^{col} = \frac{1}{2} \sum_{ij} \|y_i - y_j\|_2^2 S_{ij} \quad (14)$$

where S_{ij} is the similarity between drug d_i and d_j . By introducing the manifold regularization, the new objective function was defined as follows:

$$L_{new} = L + \frac{\mu}{2} (L_{reg}^{row} + L_{reg}^{col}) \quad (15)$$

where μ is the manifold regularization parameter. The alternating decent method was applied to minimize the above objective function to obtain the latent feature matrices X and Y . Then, the interaction probability matrix P can be calculated as follows:

$$P = XY^T \quad (16)$$

where the element P_{ij} indicates the interaction probability between drug d_i and d_j .

DDINMF

Yu et al. [96] proposed a semi-nonnegative matrix factorization method to predict enhancive and degressive DDIs (DDINMF), where enhancive (degressive) DDI refers to a drug increases (decreases) the serum concentration of itself and another drug when taken together. The authors considered m drugs with known interactions with other drugs as known drugs and n drugs without verified interactions with all known drugs as new drugs. In the training phase, different from the definition of adjacency matrix in MRMF, after extracting the enhancive and degressive DDIs information about m known drugs, the adjacency matrix A was constructed as follows:

$$A_{ij} = \begin{cases} 1, & \text{if there is an enhancive drug interaction between } d_i \text{ and } d_j \\ -1, & \text{if there is a degressive drug interaction between } d_i \text{ and } d_j \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

The authors constructed the p -dimensional feature vectors of drugs based on their chemical structure and side effects. Furthermore, the feature vectors of known drugs were combined into feature matrix F . Then, two nonnegative low-rank matrices (W and H) used to approximate matrix A could be obtained by minimizing the following objective function:

$$L = \|A - WH\|_F^2 \quad (18)$$

After obtaining matrices W and H by the method proposed by Lee et al. [97], to make the model suitable for new drugs, the authors introduced feature matrix into nonnegative matrix

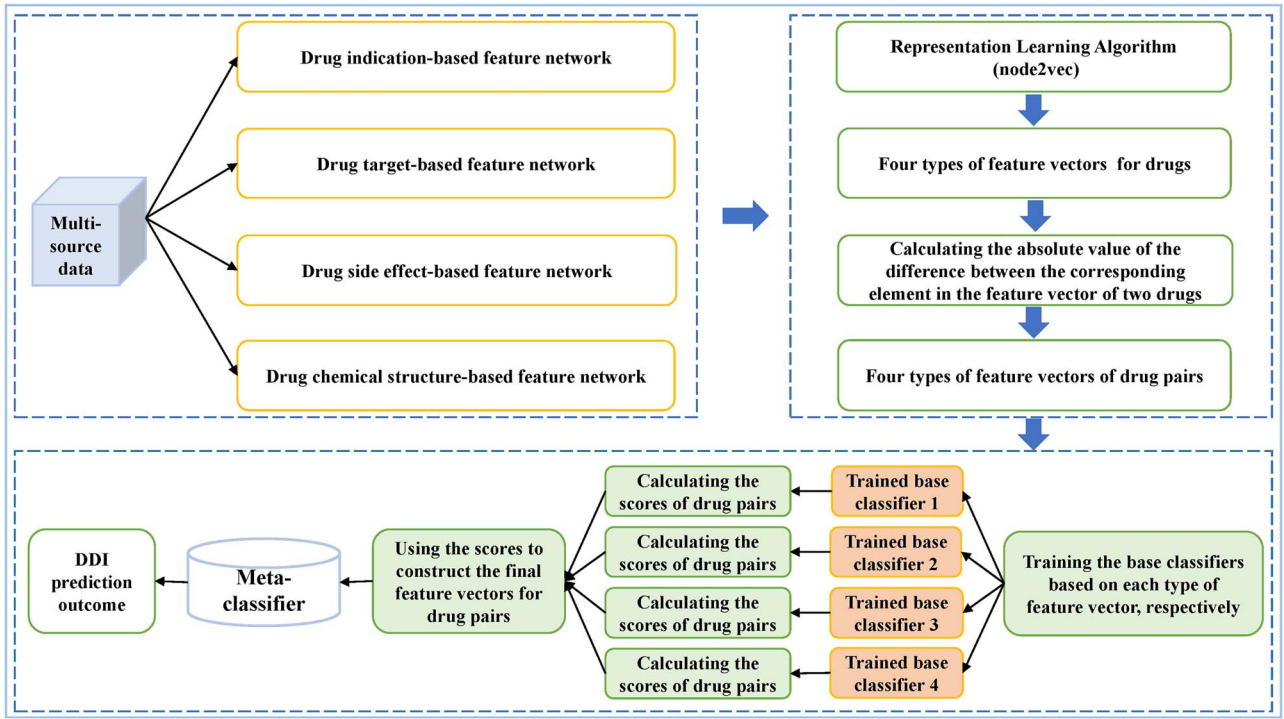


Figure 2. The flowchart of the meta-learning-based model built based on representation learning, PUL and meta-learning.

factorization. Specifically, they modeled the relationship between the feature matrix F and H as follows:

$$F \times B = H^T \quad (19)$$

where B represents the regression coefficient matrix, which was calculated by SIMPLS algorithm [98]. In the predicting phase, the feature matrix F' composed of the features of n new drugs was mapped into the latent topological space as follows:

$$H' = (F' \times B)^T \quad (20)$$

Then, the interaction probabilities between the known drugs and new drugs were calculated as follows:

$$P = (W \times H')^T \quad (21)$$

Triple matrix factorization-based unified framework

Similar to DDINMF, Shi et al. [99] presented a triple matrix factorization-based unified framework (TMFUF) to infer both enhancive and degressive DDIs. For m known drugs, the adjacency matrix A was constructed in the same way as in DDINMF, but in TMFUF, the feature vectors were constructed only based on the side effect information of drugs. Then, the author modeled the relationship between matrix A and feature matrix F as a bilinear regression, which could be represented as the triple matrix factorization:

$$A = F\Theta F^T \quad (22)$$

where Θ refers to the symmetrical projection matrix, whose role is to link the features of drugs with the interactions between drugs. To obtain the matrix Θ , the author first calculated matrix A_d^* as

follows:

$$A_d^* = \arg \min_{A_d} \|A - A_d A_d^T\|^2 \quad (23)$$

where A_d denotes the latent interaction matrix, whose row refers to the feature of corresponding drug in the latent space, and singular value decomposition was used to obtain A_d^* . Then, matrix B^* was calculated:

$$B^* = \arg \min_B \|A_d^* - FB\|^2 \quad (24)$$

where B represents the regression coefficient matrix. SIMPLS [98] was applied to solve the optimization problem to obtain B^* , and the matrix Θ was obtained as follows:

$$\Theta = B^* B^{*T} \quad (25)$$

Then, the interaction possibility matrixes could be calculated:

$$P_{n,n} = F_n \Theta F_n^T \quad (26)$$

$$P_{n,m} = F_n \Theta F^T \quad (27)$$

where F_n represents the feature matrix involving only new drugs; the elements of $P_{n,n}$ represent the interaction probabilities between new drugs, while the elements of $P_{n,m}$ refer to the interaction probabilities between new drugs and known drugs. That is, TMFUF could be used to predict not only known but also new drugs that interact with new drugs.

Local classification model via Dempster–Shafer theory of evidence

Under the assumption that similar drugs tend to interact with the same drug, Shi et al. [100] also proposed an integrated local

classification model via Dempster-Shafer theory of evidence (LCM-DS) to predict potential DDIs. The authors first constructed the drug similarity matrix through directly averaging three different drug similarity matrices (i.e. chemical structures-based, side effect-based and off-label side effect-based drug similarity matrix) derived from the work of Zhang et al. [101]. Then, based on the known DDIs as well as drug similarity, three local classification-based models (LCMs) constructed according to SVM [102], regularized least squares (RLS) [103] and multi-label K-nearest neighbors (MLKNNs) [104], were applied to calculate the interaction probabilities between drugs, respectively. Finally, the authors proposed a novel fusion method based on the Dempster-Shafer theory of evidence [105] to integrate the results of the three LCMs to obtain the final interaction probabilities between drugs.

DDIGIP

Based on the Gaussian Interaction Profile (GIP) kernel and RLS classifier, Yan et al. [106] proposed a model of DDIGIP to predict potential DDIs. Specifically, the authors first calculated eight types of drug feature vectors in the same way as MRMF, which were spliced to form the final feature vectors of drugs. Then, they used the Pearson correlation coefficient to calculate the similarities between drugs based on their respective feature vectors and then constructed the drug similarity matrix S_P . Later, to make DDIGIP also applicable to new drugs, the authors used the K-nearest neighbors (KNNs) to calculate the initial relational score between new drug d_i and known drug d_j to fill in the adjacency matrix A :

$$A(d_i, d_j) = \frac{\sum S_P(d_i, d_l) A(d_l, d_j)}{\sum S_P(d_i, d_l)}, d_l \in K_{set}^i \quad (28)$$

where K_{set}^i represents the set of top K drugs with the largest similarity to drug d_i . Next, they calculated the drug GIP similarity matrix S_G via the filled adjacency matrix. Finally, the authors used the RLS classifier to compute the predicted interaction probability matrix P as follows:

$$P' = S_G(S_G + \sigma I)^{-1} A \quad (29)$$

$$P = \frac{P' + P'^T}{2} \quad (30)$$

where I refers to the identity matrix and σ represents the regularization parameters.

Gradient boosting-based model

Qian et al. [107] proposed an extreme gradient boosting (XGBoost) classifier for the prediction of DDIs by integrating multiple features of drug pairs (Figure 3). Different from the way of constructing the feature vectors of drugs directly based on the side effect information in the previous models, the authors downloaded the data on side effects from SIDER [66], where the Unified Medical Language System (UMLS) concept IDs [108] were used as the side effects identifiers. Then, according to the dictionary MedDRA [109], they mapped the UMLS concept IDs to MedDRA concept IDs at four different levels, including preferred term (PT), high-level term (HLT), high-level group term (HLGP) and system organ class (SOC). At each level, a binary vector was constructed for each drug as its feature vector. Furthermore, the Jaccard index was used to calculate the similarities between two drugs at the four levels, respectively, which constituted the side effect-based features of corresponding drug pairs. The information about drug indications was also collected from SIDER [66] and mapped to the same four levels. Then, the indication-based features of drug

pairs were obtained in a similar way. Moreover, the authors calculated the sequence similarities between the target proteins of two drugs by Smith-Waterman algorithm [110] and then used the minimum, mean, median and maximum of the similarities between target proteins to construct the target sequence-based features of corresponding drug pairs. Similarly, the interaction scores between genes were downloaded from the study of Costanzo et al. [111], and the minimum, mean, median as well as maximum of interaction scores between the target protein gene of two drugs were used to construct genetic interaction-based features of the corresponding drug pairs. Therefore, 16 features of each drug pair could be obtained by integrating the above four types of features. Then, a feature selection method known as group minimax concave penalty (MCP) [112] was applied to obtain 11 features with significantly different value distributions between interacting drug pairs and noninteracting drug pairs, which formed the final feature vector for each drug pair. Finally, the authors applied the XGBoost classifier to calculate the interaction probability between corresponding drugs. In addition, to obtain better prediction performance, the authors optimized the hyperparameters of the classifier using the tree-structured Parzen estimator (TPE) approach [113].

Network algorithm and matrix perturbation algorithm-based model

By means of multisource data fusion, Zhang et al. [114] presented a flexible framework to integrate multiple models for DDI identification. Firstly, the authors applied the Jaccard index to calculate eight types of drug similarities based on the information about substructure, target, enzyme, transporter, pathway, indication, side effect and off side effect of drugs, respectively. Besides, according to the DDIs network constructed based on known DDIs, they computed six other kinds of drug similarities, namely, common neighbor similarity, Adamic-Adar similarity, resource allocation similarity, Katz similarity, average commute time similarity and random walk with restart similarity. Furthermore, the authors adopted two similarity-based models (constructed based on the neighbor recommender algorithm [115] as well as the random walk algorithm [116]) and one model only based on known DDIs (built according to the matrix perturbation algorithm [117]) to predict potential DDIs. Thus, according to the above-mentioned 14 types of drug similarities and known DDIs, 29 prediction models, namely, 28 similarity-based models and one model, based only on known DDIs were constructed. Finally, the weighted average ensemble rule and the classifier ensemble rule were implemented to fuse the prediction results, respectively. Specifically, the weighted average ensemble rule took the weighted average of the outputs of all prediction models, while the classifier ensemble rule took the logistic regression to map the outputs of all models to a score as the final prediction results.

Heterogeneous network-assisted inference

Cheng et al. [118] proposed a heterogeneous network-assisted inference (HNAI) framework to identify potential DDIs. The authors firstly extracted 6 946 known DDIs from DrugBank [6] as positive samples, which formed the training set with the same number of drug pairs randomly selected from drug pairs without verified interaction. Then, the authors calculated four types of drug similarities: phenotypic similarity, therapeutic similarity, chemical structural similarity and genomic similarity. Specifically, the phenotypic similarity, therapeutic similarity and chemical structural similarity were calculated according to the method proposed in the authors' previous work [119], respectively. As for genomic similarity, the authors firstly

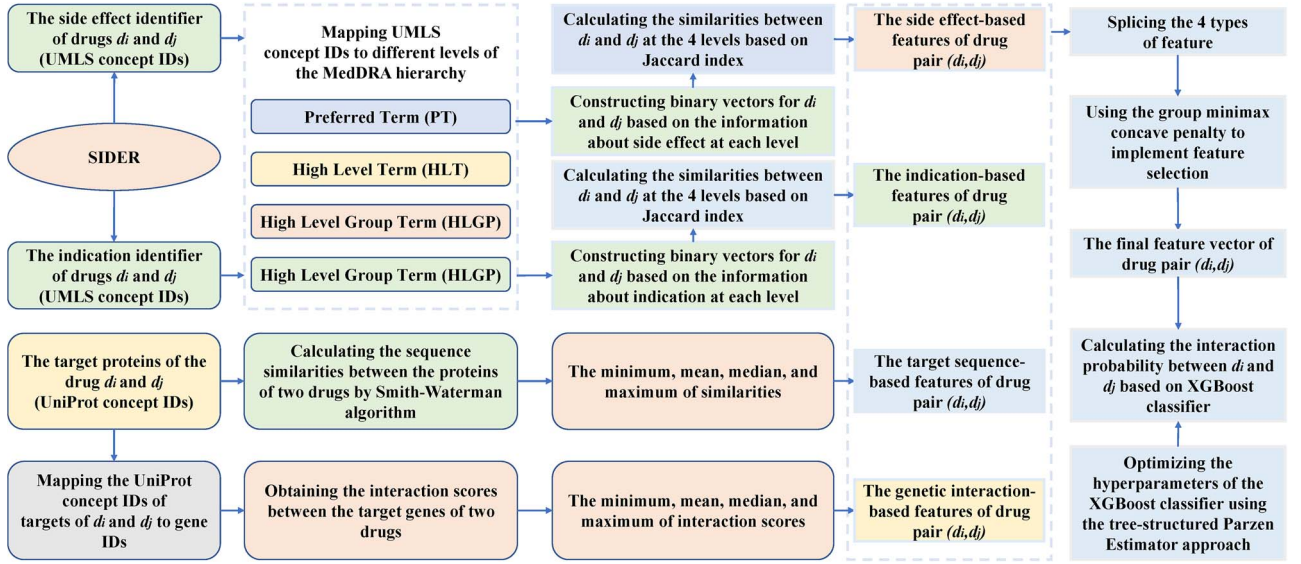


Figure 3. The flowchart of the gradient boosting-based model, where XGBoost classifier is applied to predict potential DDIs based on multiple features of drug pairs.

constructed a binary vector for each drug based on the target protein information, and then the Tanimoto coefficient [120] of the two binary vectors was regarded as the genomic similarity between the corresponding two drugs. The four types of similarity between the two drugs constituted the feature vector of the corresponding drug pair. Finally, based on the drug pairs in the training set and corresponding feature vectors, the authors trained five prediction models (namely, naive Bayes, decision tree, KNN, logistic regression and SVM) to identify potential DDIs, respectively. Moreover, the authors constructed the interaction network based on the known DDIs and carried out statistical analysis by combining the above four types of similarity. It turned out that the more similar two drugs are, the higher probability of interaction between them.

Integrated action crossing

Hunta et al. [121] proposed an integrated action crossing (IAC) method to predict the potential DDIs by focusing on the drug-enzyme and drug-transporter actions. Different from the above models where the feature vectors of drug pairs were constructed based on drug similarity, the authors proposed a method called action crossing (AC) to obtain the feature vectors of drug pairs according to the information about drug-enzyme actions [including substrate (S), inhibitor (Inh) as well as inducer (Inc)]. Specifically, for drug d_i and enzyme e_k , the action attribute vector X^{ik} was defined as follows:

$$X^{ik} = [x_S^{ik}, x_{Inh}^{ik}, x_{Inc}^{ik}] \quad (31)$$

where x_S^{ik} , x_{Inh}^{ik} and x_{Inc}^{ik} represent whether the drug d_i has a corresponding action on the enzyme e_k . Taking x_S^{ik} for an example, if drug d_i is a substrate of enzyme e_k , the value of x_S^{ik} is 1, otherwise 0. x_{Inh}^{ik} and x_{Inc}^{ik} could be obtained in a similar way. Then, the feature vector of drug pair (d_i, d_j) based on enzyme e_k was constructed:

$$F_{ij}^k = [f_1^{ijk}, f_2^{ijk}, f_3^{ijk}] \quad (32)$$

where if the p th ($p=1-3$) element of vectors X^{ik} and X^{jk} were both 1, f_p^{ijk} has a value of 1, otherwise 0. Similarly, they constructed

transport-based feature vectors of drug pairs based on the information about drug-transporter actions. The authors collected 36 enzymes as well as 35 transporters and sorted them based on their ID. Then, they calculated the feature vector based on enzyme and transporter in turn and spliced them to obtain the final feature vector. Finally, the final feature vectors of drug pairs were used to train three models (SVM, KNN and neural networks), respectively, which were used to predict potential DDIs.

SFLLN

Zhang et al. [122] proposed the sparse feature learning ensemble method with linear neighborhood regularization, named SFLLN, to predict potential DDIs. Firstly, the authors constructed the corresponding binary feature vectors for each drug based on the information about substructure, target, enzyme and pathway, respectively. Then, the same type of feature vectors for all drugs were combined into the feature matrix F_i ($i=1-4$). Secondly, the authors projected drugs from different feature spaces to the common interaction space by approximating the interaction probability matrix P with the product of the feature matrix F_i and the nonnegative projection matrix G_i . Besides, they controlled the sparsity of projection matrixes through minimizing $\sum_{k=1}^{n_i} \|G_i[:, k]\|^2$ to improve the generalization ability of the model, where n_i referred to the number of columns in matrix G_i . Therefore, they defined the objective function as follows:

$$\min_{G, P} \|A - P\|_F^2 + \mu \sum_{i=1}^4 \|F_i G_i - P\|_F^2 + \lambda \sum_{i=1}^4 \sum_{k=1}^{n_i} \|G_i[:, k]\|^2 \quad (33)$$

where A refers to the adjacency matrix constructed based on known DDIs; λ and μ refer to free parameters. Besides, by assuming that the predicted DDIs have the same structure as the known DDIs, the authors defined the Lagrangian function as follows to extract the data structure from known DDIs:

$$L = \frac{1}{2} \|A - (C \otimes W) A\|_F^2 + \frac{\mu}{2} \|(C \otimes W) e\|_F^2 - \lambda ((C \otimes W) e - e) - \text{tr}(\Phi W) \quad (34)$$

where \otimes denotes the Hadamard product; C is an indicator matrix, where $C[i,j]=0$ if $i=j$, otherwise $C[i,j]=1$; e is a n_d -dimensional vector with all elements being 1; n_d is the total number of drugs; Φ refers to the Lagrange multiplier; matrix W reflects the intrinsic structure of known DDIs. Then, W was obtained by taking the derivative of L with respect to W and setting the derivative to 0. In order to ensure that the drugs retain their internal structure after projection, matrixes P and W should meet the following requirement:

$$P[i,j] - \sum_{k=1}^{n_d} W[i,k] P[k,j] \approx 0 \quad (35)$$

Therefore, by algebraic transformation, the authors defined the linear neighborhood regularization [123]:

$$\sum_{i=1}^{n_d} \sum_{j=1}^{n_d} P[i,j] \left(P[i,j] - \sum_{k=1}^{n_d} W[i,k] P[k,j] \right) = \text{tr} (P^T (I - W) P) \quad (36)$$

By combining Formulas (33) and (36), the authors defined the final objective function:

$$\min_{G,P} \|A - P\|_F^2 + \mu \sum_{i=1}^4 \|F_i G_i - P\|_F^2 + \delta \text{tr} (P^T (I - W) P) + \lambda \sum_{i=1}^4 \sum_{k=1}^{n_d} \|G_i[:,k]\|_1^2 \quad (37)$$

The authors set the partial derivative of L to P as 0 to obtain the relationship between P and G :

$$P = (\delta (I - W) + (1 + m\mu) I)^{-1} \left(A + \mu \sum_{k=1}^4 F_k G_k \right) \quad (38)$$

Then, Formula (37) could be rewritten as follows:

$$\min_G \sum_{k=1}^4 \text{tr} \left(\mu \left(G_k^T F_k^T F_k G_k - 2\mu A^T (\delta (I - W) + (1 + m\mu) I)^{-1} F_k G_k + \lambda G_k^T E_k G_k \right) - \mu^2 \sum_{k_1=1}^4 \sum_{k_2=1}^4 \text{tr} \left(G_{k_1}^T F_{k_1}^T \left((\delta (I - W) + (1 + m\mu) I)^{-1} \right)^T F_{k_2} G_{k_2} \right) \right) \quad (39)$$

where all elements in matrix $E_k \in R^{n_k \times n_k}$ are equal to 1 and n_k represents the number of columns in F_k . Finally, the authors solved Formula (39) by semi-nonnegative matrix factorization algorithm to get G_k and then obtained the interaction probability matrix P based on Formula (38).

Deep learning-based models

Given the successful application of deep learning algorithms in the fields such as natural language processing (NLP) [124, 125] and pattern recognition [126, 127] in recent years, more and more researchers have built models based on deep learning methods to predict potential DDIs. Deep learning-based models can not only automatically extract the features of drugs but also effectively integrate the features of drugs by multiple modules to obtain the features of the corresponding drug pairs. In addition, NLP-based models could be applied to mine a large number of DDIs from the literatures. However, as with the models constructed based on the traditional machine learning algorithms, the scarcity of reliable negative samples severely limits the performance of deep learning-based models. Besides, because the goal of training is to obtain the optimal values of the parameters and the most significant features of the drug pairs, using deep learning-based models usually takes more time to make predictions. We have a brief introduction to some of them below.

DDIMDL

Deng et al. [128] developed a multimodal deep learning framework to identify unknown DDIs (DDIMDL). Firstly, the authors constructed the corresponding binary feature vector for each drug based on the information about substructure, target, enzyme and pathway, respectively. Secondly, the Jaccard index was used to calculate the similarities between drugs and four corresponding similarity matrices (i.e. S_s , S_t , S_e and S_p) were constructed later, where the i th row of each similarity matrix was regarded as the corresponding feature vector of drug d_i . Thirdly, for each similarity matrix, its i th and j th rows were fed into the DNN to calculate the interaction probability between d_i and d_j . Finally, the authors took the average of the interaction probabilities obtained based on the four similarity matrices as the final result.

Substructure-substructure interaction for drug-drug interaction

Considering that DDIs were caused by the interactions between the substructures of the corresponding two drugs, Shi et al. [129] developed a deep learning-based model named substructure-substructure interaction for drug-drug interaction (SSI-DDI) to predict potential DDIs (Figure 4). Firstly, for drug d_i , based on the software RDKit (<https://www.rdkit.org/>), the SMILES string of d_i download from DrugBank [6] was converted into a molecular graph, where nodes represented the atoms contained in d_i and edges referred to the bonds between corresponding atoms. Secondly, they built a module made up of four graph attention (GAT) layers in series (layer 1 → layer 2 → layer 3 → layer 4), where layer 1 was used to extract the feature vector (64-dimension) of each node based on the molecular graph, while the other three layers were applied to update the vector output from the previous layer. Then, by integrating the feature vectors of all atom nodes in the molecular graph of d_i , the vector recording the substructure information extracted by the k th layer could be obtained:

$$g_i^k = \sum_{p=1}^n \beta_p v_i^{k,p} \quad (40)$$

where n is the total number of nodes; parameter β_p indicates the importance of the p th node, which could be obtained by SAGPooling [130]; $v_i^{k,p}$ refers to the feature vector of the p th node extracted by the k th layer. For the substructure of d_i extracted by the k_1 th layer and substructure of d_j extracted by the k_2 th layer, the co-attention mechanism was applied to calculate the importance score $r_{k_1 k_2}$ of the interaction between these two substructures to the final DDI prediction. Finally, the interaction probability between d_i and d_j could be obtained by integrating the interactions between substructures of two drugs:

$$P_{ij} = \sigma \left(\sum_{k_1=1}^4 \sum_{k_2=1}^4 r_{k_1 k_2} g_i^{k_1} M g_j^{k_2} \right) \quad (41)$$

where σ refers to the sigmoid function; M is a learnable matrix, which was obtained by training SSI-DDI on 1024 known DDIs using the Adam optimizer [131].

Substructure-aware tensor neural network model for DDI prediction

Shi et al. [132] also constructed another model named substructure-aware tensor neural network model for DDI prediction (STNN-DDI), which could be used to predict the types of DDI. Firstly, each drug was represented by an n -dimensional binary vector, where

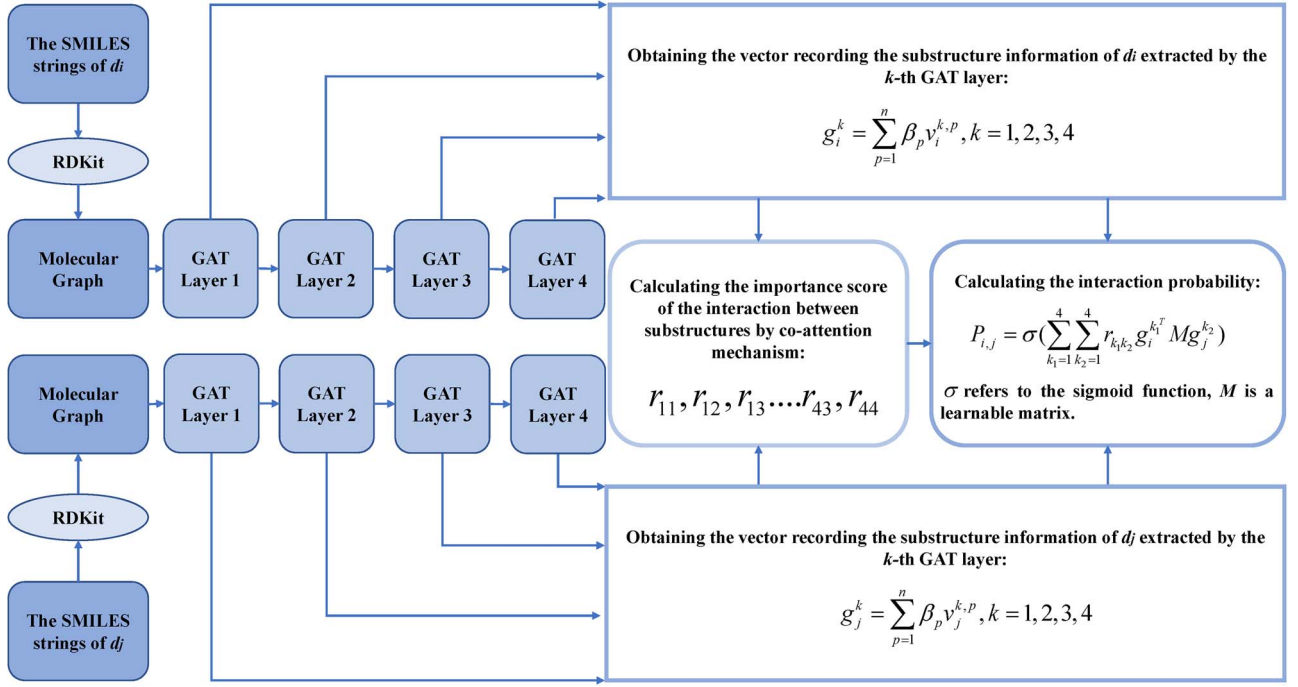


Figure 4. The flowchart of SSI-DDI, which is applied to predict potential DDIs based on the interactions between drug substructures.

n represents the total number of substructures under study. For example, drug d_i could be denoted by the vector $e_i = [e_1^i, e_2^i, \dots, e_n^i]$, where the value of e_j^i is 1 if d_i contains the j th substructure and 0 otherwise. Then, for drugs d_i and d_j , the authors used P_{ijk}^d to represent the probability of the k th type of interaction between d_i and d_j , which could be defined as follows:

$$P_{ijk}^d = \sum_{p=1}^n \sum_{q=1}^n P_{pqk}^s e_p^i e_q^j \quad (42)$$

where P_{pqk}^s refers to the probability that there is k th type of interaction between the p th substructure and the q th substructure. In order to get the interaction probabilities between substructures, the authors constructed a tensor named ST , where both axes x and y represented the substructures, while the axis z referred to the DDI types, and set $ST_{pqk} = P_{pqk}^s$. Therefore, Formula (42) could be rewritten as

$$P_{ijk}^d = \sum_{p=1}^n \sum_{q=1}^n ST_{pqk} e_p^i e_q^j \quad (43)$$

Based on the CP decomposition [133], the tensor ST could be approximated by three factor matrices A , B and C :

$$ST \approx \sum_{z=1}^r \lambda_z (a_z \times b_z \times c_z) = \langle \lambda; A, B, C \rangle \quad (44)$$

where λ denotes the r -dimensional vector used to normalize the columns of three factor matrices; $A \in R^{n \times r}$, $B \in R^{n \times r}$ and $C \in R^{f \times r}$ record the latent information of the x , y and z axes, respectively; r is the number of rank-one tensors decomposed by ST ; f represents the total number of DDI types. Besides, by assuming that the rows of matrix A and B represent the embedding of the corresponding substructure, while the rows of C refer to the embedding of the corresponding DDI type, the authors introduced

the multi-linear tensor transformation [134] to define the interaction probability P_{ijk}^d of the k th type of interaction between d_i and d_j :

$$P_{ijk}^d = \sum_{z=1}^r \lambda_z (a_z \times b_z \times c_z) \bar{\times}_3 v_k \bar{\times}_2 e_i \bar{\times}_1 e_j + \text{bias} \\ = [(e_i^T \times A) \otimes (e_j^T \times B) \otimes (v_k^T \times C)] \lambda + \text{bias} \quad (45)$$

where $\bar{\times}_n$ refers to the mode- n product; v_k represents the one-hot vector encoding the k th type of interaction; parameter bias is added to enhance the robustness of STNN-DDI. Then, they defined the loss function F :

$$F = \sum_{(d_i, d_j) \in D_{\text{train}}} (PD_{ijk} - P_{ijk}^{d*})^2 \quad (46)$$

where D_{train} is the training set consisting of all positive samples and an equal number of randomly selected negative samples; the value of PD_{ijk} is 1 if there is the k th type of interaction between drug d_i and d_j and 0 otherwise. Finally, the authors constructed a fully connected neural network model based on Formulas (45) and (46) to obtain matrices A , B , C , λ and bias according to the training set. Then, the interaction probability between drugs could be calculated based on Formula (45).

META-DDIE

Deng et al. [135] proposed a few-shot computational model named META-DDIE to predict the types of DDIs, which consisted of a representation module and a comparing module. In the representation module, the authors first constructed a binary vector S_i for drug d_i based on its structure information. For drug pair d_i - d_j , its feature vector $F_{i,j}$ was defined as follows:

$$F_{i,j}[k] = \begin{cases} 0, & \text{if } S_i[k] = 0 \text{ and } S_j[k] = 0 \\ 1, & \text{otherwise} \end{cases} \quad (47)$$

Then, the authors employed a neural network to encode the vector F_{ij} to a embedding vector E_{ij}^1 , and applied another neural network to decode a new feature vector F'_{ij} from vector E_{ij}^1 . To train the framework consisting of encoder and decoder, the authors defined the loss function:

$$L_{ij} = \sum_{k=1}^n (F_{ij}[k] \log(F'_{ij}[k]) + (1 - F_{ij}[k]) \log(1 - F'_{ij}[k])) \quad (48)$$

where n is the dimension of the vector F_{ij} . Secondly, based on the SMILES of drugs, the authors applied a chemical sequential pattern mining (SPM) algorithm [136] to obtain a set of discrete frequent substructures of drugs in the database. The k th frequent substructure (denoted by a single-hot vector v_k) was fed into the above neural network for encoding to obtain corresponding embedding vector E_k^2 . Then, the vector E_{ij}^1 could be projected on a subspace defined by span $([E_1^2, E_2^2, \dots, E_{n_s}^2])$:

$$E_{ij}^1 = E_1^2 r_{ij}^1 + E_2^2 r_{ij}^2 + \dots + E_{n_s}^2 r_{ij}^{n_s} \quad (49)$$

where r_{ij}^k ($k=1,2,\dots,n_s$) represents the projection coefficient, which could be calculated via the method proposed by Huang et al. [137]; n_s denotes the total number of frequent substructures. The vector $r_{ij} = [r_{ij}^1, r_{ij}^2, \dots, r_{ij}^{n_s}]$ was regarded as the final representation the drug pair d_i-d_j . For the few-shot learning, the authors divided drug pairs into training sets and test sets, and both of them were further divided into support set as well as query set. For drug pairs DP_p in the support set and DP_q in the query set, their representation were fed into the comparing module constructed as in the study [138] to obtain an n_t -dimensional similarity vector $S_{p,q}$ between the two drug pairs, where n_t refers to the number of DDI types. Then, they defined the loss function based on mean square error to train the model:

$$L = \sum_{p=1}^{n_p} \sum_{q=1}^{n_q} (\max(S_{p,q}) - l_{p,q})^2 \quad (50)$$

where n_p and n_q represent the number of drug pairs in the support set and the query set, respectively. If DP_p and DP_q have the same type of DDIs, the value of $l_{p,q}$ is 1, otherwise 0. Then, the model was trained by minimizing the loss function L . Finally, the authors applied the trained model to calculate the similarity vector $S_{x,y}$ between drug pairs DP_x in the support set and DP_y in the query set. The DDIs type corresponding to the maximum value in the vector $S_{x,y}$ was regarded as the type of drug pair DP_y .

Deep attention neural network-based drug-drug interaction prediction model

Liu et al. [139] developed a deep attention neural network-based drug-drug interaction prediction model (DANN-DDI) to identify potential DDIs. Specifically, the authors first constructed five networks, including the drug-substructure network, drug-target network, drug-enzyme network, drug-pathway network and DDI network. For drug d_i , the authors applied structural deep network embedding method [140, 141] to learn its embeddings (i.e. $E_i^s, E_i^t, E_i^e, E_i^p$ and E_i^d) from the above networks, respectively. Then, the authors constructed the comprehensive vector $E_i = [E_i^s, E_i^t, E_i^e, E_i^p, E_i^d]$ of d_i based on the above embeddings. For drug d_i and d_j , their comprehensive vectors were used as the input of the attention neural network [142] to obtain the feature vector of drug pair d_i-d_j . Finally, the feature vector was fed into the

framework consisting of the input layer, multiple fully connected hidden layers and the output layer, and the softmax function was applied to calculate the interaction probability between d_i and d_j based on the output of the framework.

Multi-relational contrastive learning graph neural network

Xiong et al. [143] proposed a model named multi-relational contrastive learning graph neural network (MRCGNN) to predict the types of DDIs. Firstly, by taking drugs as nodes and known DDIs as edges, the authors constructed a multi-relational DDI event graph $G=(V,E,T)$, where V and E represent the set of all drug nodes and edges, respectively, T denotes the set of all DDIs types. Secondly, after obtaining the molecular graph of each drug in the same way as used in SSI-DDI, the authors utilized TrimNet [144] to extract the features of drugs based on the corresponding molecular graph and constructed feature matrix F by combining the features of all drugs. Then, the authors employed the relational graph convolutional network (R-GCN) encoder [145] to learn the original representation vectors of drugs from the graph G with the features of drugs as node attributes, and the representations of all drugs formed the matrix H . Besides, the global representation g was defined:

$$g = \Gamma(H) \quad (51)$$

where Γ refers to the readout function. Thirdly, in order to implement the multi-relational contrastive learning on G , the authors corrupted the graph G by shuffling the features of drug nodes and edges to obtain corrupted graphs $G_v=(\tilde{V},E,T)$ and $G_e=(V,\tilde{E},T)$, which were fed into the R-GCN encoder to obtain the corresponding drug representation matrix H_v and H_e , respectively. Given that the training goal of contrastive learning was to maximize the consistency between H and g , as well as the difference between H_v/H_e and g , the authors defined two loss functions:

$$L_v = -\frac{1}{|V|+|\tilde{V}|} \left(\sum_{i=1}^{n_d} E_{(V,E,T)} [\log D(H[i],g)] + \sum_{i=1}^{n_d} E_{(\tilde{V},E,T)} [\log(1-D(H_v[i],g))] \right) \quad (52)$$

$$L_e = -\frac{1}{|V|+|\tilde{V}|} \left(\sum_{i=1}^{n_d} E_{(V,E,T)} [\log D(H[i],g)] + \sum_{i=1}^{n_d} E_{(V,\tilde{E},T)} [\log(1-D(H_e[i],g))] \right) \quad (53)$$

where $D(H[i],g) = \sigma H[i] W g$; W represents a trainable parameter matrix; n_d refers to the total number of drugs. For drug pair d_i-d_j , the authors spliced the corresponding features and representations (i.e. $F[i], F[j], H[i]$ and $H[j]$) together to obtain the final representations r_{ij} of the drug pair, which was fed into the multilayer perceptron (MLP) followed by a Softmax function to implement multi-class prediction:

$$P_{ij} = \text{soft max}(MLP(r_{ij})) \quad (54)$$

where P_{ij} is a n_t -dimensional vector; n_t is the number of DDIs types; $P_{ij}[k]$ represents the probability of the k th type of interaction between drug d_i and d_j . Then, the authors defined another loss function:

$$L_c = -\sum_{(d_i,d_j) \in \Omega} \sum_{k=1}^{n_t} L_{ij}^k \log P_{ij}[k] \quad (55)$$

where Ω represents the training set; L_{ij}^k is the true label of drug pair d_i - d_j , if there is the k th type of interaction between d_i and d_j , L_{ij}^k has the value 1, otherwise 0. To train MRCGNN, the authors defined the final loss function based on the above three loss functions:

$$L = L_c + \alpha L_v + \beta L_e \quad (56)$$

where α and β refer to the hyperparameters used to balance different loss functions.

Multichannel feature fusion model for multi-typed DDI prediction

Chen et al. [146] developed a multichannel feature fusion model for multi-typed DDI prediction (MCFF-MTDDI), which consisted of three modules, namely, feature extraction module, feature fusion module and classifier module (Figure 5). The authors firstly removed all <DRUGBANK::ddi-interactor-in::Compound::Compound> edges to remove the DDI information from the drug repurposing knowledge graph (DRKG) [147]. The remaining triples made up the biomedical knowledge graph (KG) dataset after removing the isolated drug nodes. In the feature extraction module, the authors extracted three types of KG representations (namely, initial embedding representation, subgraph mean representation and subgraph frequency representation) for each drug through corresponding methods based on the KG dataset, respectively. Besides, they also obtained the Morgan fingerprint vector [148] of each drug through RDKit (<https://www.rdkit.org/>) based on the SMILES string of corresponding drug. Then, the Morgan fingerprint vectors of the two drugs were spliced together to construct the chemical structure-based feature of the corresponding drug pair. Moreover, the authors constructed the extra label-based feature of each drug pair for multi-label prediction. Specifically, after removing the drugs without SMILES, there were 12 362 different DDI types in the dataset, corresponding to 12 362 labels. Then, from the DDI labels involving more than 10 000 drug pairs, the authors selected 200 labels involving the fewest drug pairs to build the target label set, and the remaining 12 162 labels made up the extra label set. For the drug pair (d_i, d_j) , a 12 162D binary vector $H_{EL(d_i, d_j)} = [h_{ij}^1, h_{ij}^2, \dots, h_{ij}^{12162}]$ was constructed as its extra label vector, where if there was k th type of interaction between d_i and d_j , the value of h_{ij}^k is 1, otherwise 0. Then, principal component analysis (PCA) was applied to reduce the dimension of extra label vector to obtain vector $H'_{EL(d_i, d_j)} \in R^{300}$, and the extra label-based feature vector was defined:

$$F_{EL(d_i, d_j)} = \text{ReLU}(WH'_{EL(d_i, d_j)} + b) \quad (57)$$

where W represents the trainable weight and b refers to the bias. In the feature fusion module, the state encoder consisting of two fully connected layers and two state vector strategy blocks was used to integrate the KG representations to obtain the KG fusion representations of the drug pair (d_i, d_j) . Moreover, the chemical structure-based feature and KG fusion representations were input into a GRU-based multichannel feature fusion framework to obtain the fused feature vector $F_{FU(d_i, d_j)}$ of the drug pair (d_i, d_j) . In the classifier module, the vector $F_{FU(d_i, d_j)}$ was used as the input of the multi-class classifier to implement the multi-class classification tasks, while the extra label-based feature vector $F_{EL(d_i, d_j)}$ and vector $F_{FU(d_i, d_j)}$ were concatenated and input into the multi-label classifier to implement the multi-label classification. It should be pointed out that both classifiers consisted of two

fully connected layers, where the number of neurons in the last layer was equal to the number of DDI types of the corresponding classification task.

DSIL-DDI

Tang et al. [149] proposed a model called DSIL-DDI to predict potential DDIs by implementing causal representation learning [150] on the substructures of drugs. Specifically, for drug d_i , the authors first constructed the molecular graph for d_i in the same way as used in SSI-DDI, which was input into graph neural networks (GNNs) to obtain its substructure representations, where the p th substructure was represented by vector S_i^p . After obtaining the representations of two substructures, the priori representation of the interaction between them was defined:

$$ssi_{pq} = W_{SSI} S_i^p \otimes S_j^q \quad (58)$$

where W_{SSI} represents the learnable weight matrix. Then, the attention weights were used to modify the priori representation to obtain the posteriori representation:

$$postssi_{pq} = \text{MLP}(\text{concat}(S_i^p, S_j^q)) \times ssi_{pq} \quad (59)$$

where MLP denotes the multilayer perceptron and concat refers to the concatenate operation. In a similar way, the authors calculated the posteriori representations between all substructures of d_i and all substructures of d_j , which were integrated to obtain the substructure interaction matrix, where the row and column represented the substructure of two drugs, respectively, and the element referred to the posteriori representation of corresponding two substructures. Finally, the substructure interaction matrix was input into the single-layer linear network to obtain the interaction probability between d_i and d_j .

DSN-DDI

Based on the intra-view and inter-view representation learning methods, Li et al. [151] developed a novel model named DSN-DDI to identify potential DDIs. Firstly, the authors obtained the molecule graph of each drug in the same way as used in SSI-DDI. Besides, for the drug pair (d_i, d_j) , they built a bipartite graph by connecting each atom node in the molecule graph of d_i with all nodes in the molecule graph of d_j in turn. Secondly, to learn the representations of nodes in the graphs, the authors constructed four identical DSN encoders (composed of the representation extraction layer, the intra-view layer and the inter-view layer), which formed the DSN encoder module in series. Specifically, in the first DSN encoder, the molecular graph of drug d_i was input into the representation extraction layer to obtain the representation of each node. Then, in the intra-view layer, the GAT [152] was applied to update the node representation by capturing the interactions between the atoms of d_i . Moreover, the node representations of drug d_j were extracted and updated in a similar way. Besides, the bipartite graph was fed into the representation extraction layer to extract the representation of each node. Then, in the inter-view layer, the node representation of two drugs were updated by capturing the interactions between the atoms of two drugs via the co-attentional mechanism. The other three DSN encoders were applied to update the output from the previous encoder. Then, the output of the DSN encoder module was input into the self-attention graph (SAG) pooling layer to learn the drug representations and obtain the embedding vectors of d_i and d_j . Finally, the co-attention scoring function was used to calculate

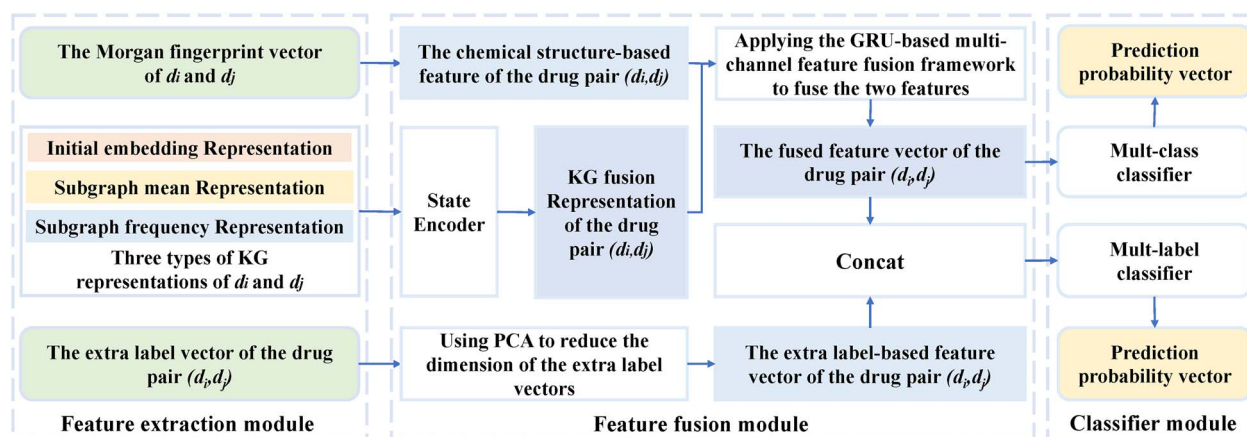


Figure 5. The flowchart of MCFF-MTDDI, which consists of three modules: feature extraction module, feature fusion module and classifier module.

the interaction probability of corresponding drug pair based on the embedding vectors.

BioDKG-DDI

Based on the self-attention mechanism [153], Ren et al. [154] constructed a model of BioDKG-DDI to predict potential DDIs (Figure 6). Firstly, based on the molecular structure information of drugs recorded in DrugBank [6], a novel molecular representation method named Mol2Context-vec [155] was used to extract the molecular structure features of drugs. Secondly, according to four types of association information (namely, drug-carrier association, drug-enzyme association, drug-target association and drug-transporter association) recorded in DrugBank, they constructed the drug knowledge graph (DKG), where nodes represented biological entities and edges referred to corresponding associations. Then, ComplEx-DURA [156] was applied to extract the global association features of drugs based on the DKG. Thirdly, according to the association information of drug-carrier, the author constructed an adjacency matrix, where rows and columns represent drugs and carriers, respectively. The Euclidean distance between i th row and j th row of the adjacency matrix was calculated as the similarity between drug d_i and d_j . Then, the similarity matrix based on the association information of drug-carrier was obtained. Besides, the corresponding drug similarity matrices were obtained based on the other three types of associations in a similar way. Then, the similarity network fusion method [157] was used to integrate four similarity matrices to get the final similarity matrix, where each row was regarded as the similarity features of the corresponding drug. Finally, the author used the self-attention mechanism to integrate the above three types of features of each drug to get the final feature and input the final features of the two drugs into the deep neural networks (DNNs) to obtain the interaction probability of the corresponding drug pair.

MDF-SA-DDI

According to multisource feature fusion and self-attention mechanism [158], Lin et al. [159] developed a model named MDF-SA-DDI to predict DDIs (Figure 7). Firstly, for each drug, three binary vectors were constructed based on the information about targets, enzymes and chemical structures of drugs, respectively. Then, based on each type of binary vector, the author calculated the similarity between drugs by Jaccard index and constructed the corresponding similarity matrix. The i th row of the three similarity matrices were spliced as the feature vector of drug d_i . Secondly,

the authors used the Siamese network [160], convolutional neural networks (CNNs) and autoencoders with self-attention mechanism to fuse the feature vectors of two drugs to obtain the feature vectors of drug pair (d_i, d_j) , respectively. Finally, the multi-head self-attention mechanism was applied to integrate the above three types of feature vectors of drug pair (d_i, d_j) to obtain the final feature vector, which was input into the full connection layer to calculate the interaction probability between d_i and d_j .

Deep feed-forward network-based model

Lee et al. [161] proposed a novel DDI prediction model based on autoencoders and the deep feed-forward network. Firstly, the authors calculated three types of drug similarity. Taking drug d_i and d_j as an example, a binary vector was constructed based on the substructure information of each drug, and then the Tanimoto coefficient of the two vectors was calculated as the substructure-based similarity between d_i and d_j . Besides, based on the target genes of drugs, the authors calculated the target gene-based similarity according to the functional interaction (FI) network downloaded from BioGrid [162]:

$$TS_{ij} = \frac{\left| \left\{ (x, y) \mid x \in G_i, y \in G_j, d(x, y) \leq t_i \right\} \right|}{\left| \left\{ (x, y) \mid x \in G_i, y \in G_j \right\} \right|} \quad (60)$$

$$t_i = \max \{d(x, y) \mid x, y \in G_i\} \quad (61)$$

where G_i and G_j represent the set of target genes of drug d_i and d_j , respectively, (x, y) refers to the gene pair composed of gene x and y , $d(x, y)$ is the distance between x and y in the FI network. In a similar way, the GO term-based drug similarity could be calculated according to the GO term and GO graph [163]. Then, the corresponding similarity matrixes were constructed based on the three types of similarity, respectively. Secondly, for the drug pair (d_i, d_j) , the i th and j th rows of each similarity matrix were input into the autoencoder to obtain the feature vector of (d_i, d_j) . Finally, the three feature vectors of the drug pair were spliced to obtain the final feature vector, which was input into the deep feed-forward network to obtain the interaction probability between d_i and d_j .

R²-DDI

Lin et al. [164] developed a model relation-aware feature refinement for DDI prediction (R²-DDI) to predict potential DDIs. Specifically, for drug d_i , the authors first constructed its molecular

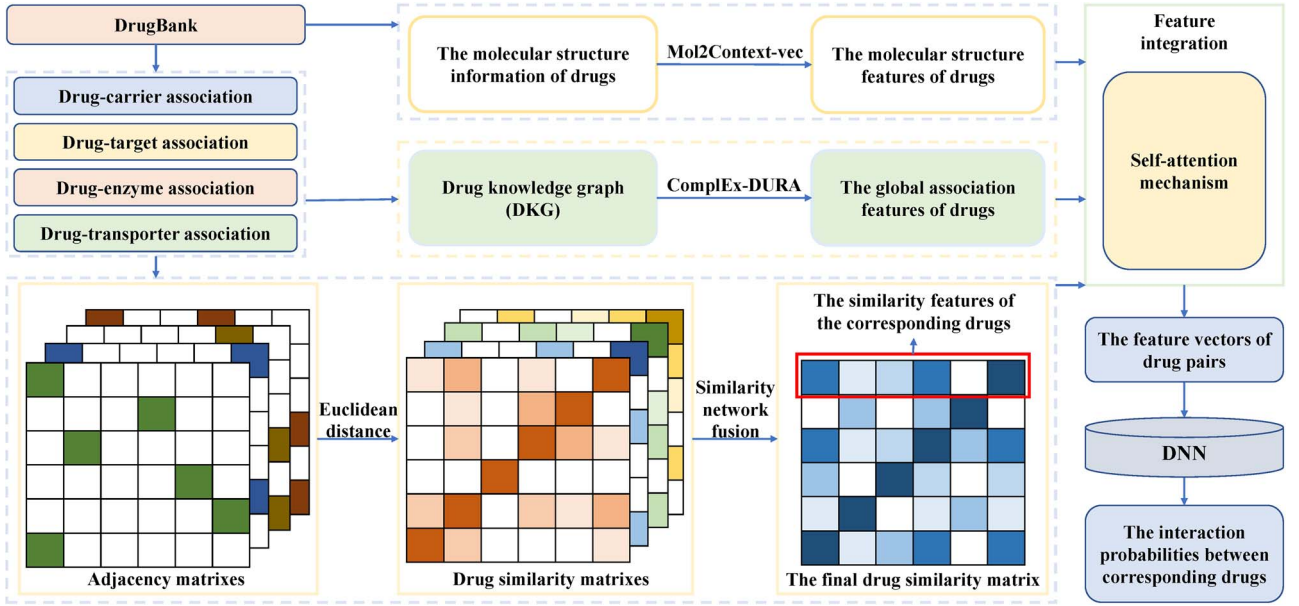


Figure 6. The flowchart of BioDKG-DDI, where the self-attention mechanism is used to fuse the features of two drugs to obtain the features of the corresponding drug pair, which is input into the DNN to obtain the corresponding interaction probability.

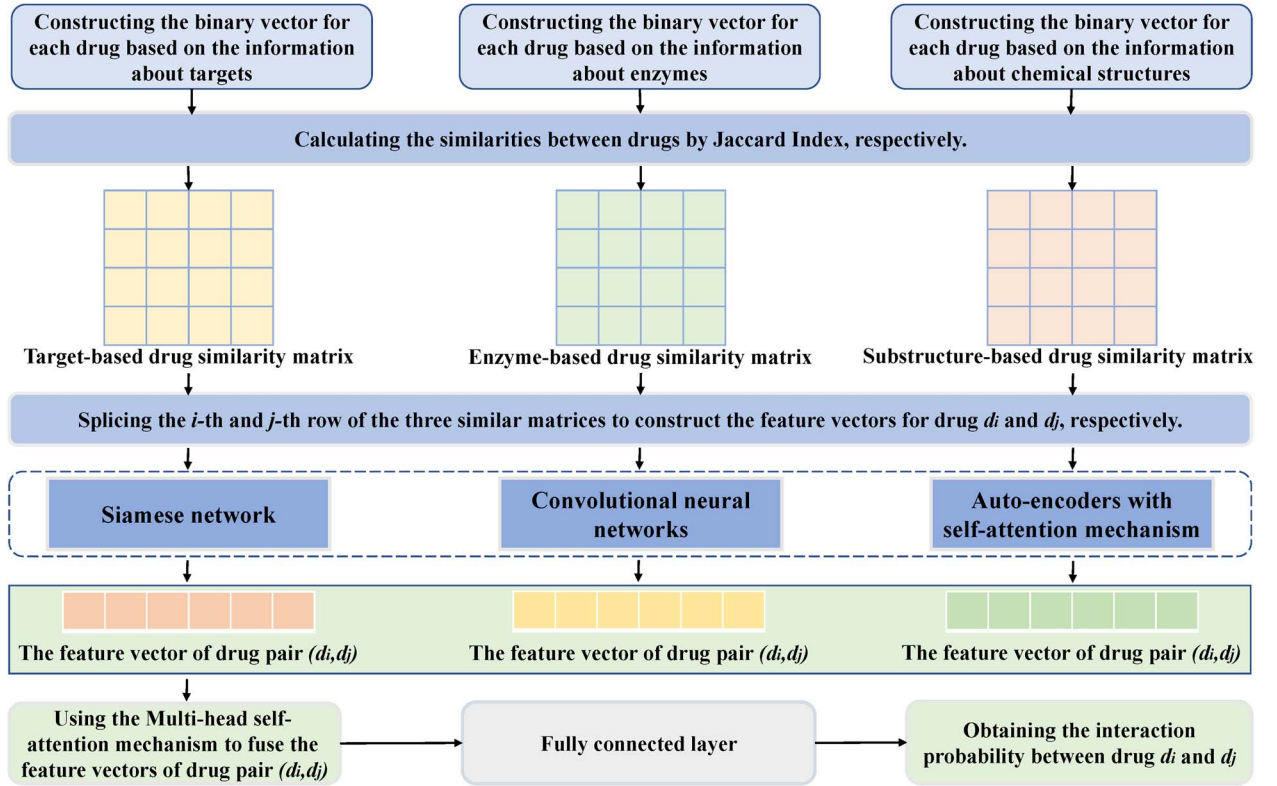


Figure 7. The flowchart of MDF-SA-DDI constructed based on multisource feature fusion and self-attention mechanism.

graph in the same way as used in SSI-DDI, which was input into DeeperGCN to obtain the graph features vector E_i . The k th type of interaction was represented by a learnable vector $T_k \in \mathbb{R}^d$, where d represents the dimensions of the interaction feature. Then, in order to construct the relationship among E_i , E_j and T_k , the authors calculated the refinement vectors based on the MLP, which were added to the corresponding original feature vectors to obtain the

refined features:

$$E'_i = E_i + \text{MLP}(\text{concat}(E_i, E_j, T_k)) \quad (62)$$

$$E'_j = E_j + \text{MLP}(\text{concat}(E_i, E_j, T_k)) \quad (63)$$

$$T'_k = T_k + \text{MLP}(\text{concat}(E_i, E_j, T_k)) \quad (64)$$

Finally, the probability of the k th types of interaction between d_i and d_j was calculated as follows:

$$P_{ij}^k = \text{Sigmoid} \left(T_k^T \text{MLP} \left(\text{concat} \left(E_i', E_j' \right) \right) \right) \quad (65)$$

Graph kernel-based approach

In view of the successful application of NLP in biomedicine [165] and computational biology [166], Zhang et al. [167] proposed a novel model to detect rapidly accumulating PK DDIs from the biomedical literatures. Unlike the above models of implementing predictions based on the database recording known DDIs, the PK DDI corpus built by Wu et al. [168] were employed in this study, which recorded 428 abstracts derived from literatures on PK DDIs. The drug pairs consisting of two drugs appearing in the same sentence were considered as candidate samples. Besides, each sentence with candidate samples was represented by a dependency graph (constructed based on the syntactic structure of sentence) [169] as well as a shallow semantic graph (built according to the shallow semantic relation structure of sentence) [170], respectively. Then, according to the method proposed by Airola et al. [171], the authors constructed all-path graph kernels to describe the connections between syntactic and semantic within the sentences. Finally, the graph kernels were used to train the least squares SVM classifier [171], which was applied to identify potential PK DDIs from the literatures.

Semantic predication-based model

Based on two widely used NLP tools: MetaMap [172] and SemRep [173], Zhang et al. [174] proposed a method to identify potential DDIs via semantic predications. Specifically, firstly, the authors extracted the drug list from clinical data and used MetaMap to map them to the concepts in UMLS [108]. Secondly, from SemMedDB (a database composed of semantic predications generated by SemRep) [175], they extracted four types of semantic predication (namely, drug–predicate–biological function, gene–predicate–biological function, gene–predicate–drug and drug–predicate–gene), where each semantic predication referred to a subject–predicate–object triplet with the UMLS concept as subject and object as well as semantic relationships from the UMLS semantic network as predicates. Thirdly, gene names were normalized to approved gene symbols based on Gene Nomenclature Committee dataset [176]. Fourthly, according to two types of pathway schemas, all drug–drug pairs based on the combinations of semantic predications were collected. In the first schema, drug d_i affects drug d_j through acting on gene g_k (i.e. $d_i \rightarrow g_k \rightarrow d_j$). In the second schema, d_i affects g_{k1} , while d_j affects g_{k2} , where both g_{k1} and g_{k2} regulate the same biological function (i.e. $d_i \rightarrow g_{k1} \rightarrow \text{biological function} \leftarrow g_{k2} \leftarrow d_j$). Finally, the predicted potential DDIs were obtained by filtering out known DDIs from the collected drug–drug pairs.

Att-BLSTM

Zheng et al. [177] proposed a model named Att-BLSTM to extract DDIs from the biomedical literatures by combining attention mechanism and the recurrent neural network (RNN) with bidirectional long short-term memory (BLSTM). The network architecture of Att-BLSTM was made up of six components, namely, the input layer, embedding layer, input attention layer, merging layer, BLSTM layer and softmax layer. Besides, the DDI-2013 corpus [178] was used in this study, which consisted of the texts describing drugs, and the drug pair in each sentence were manually

labeled as either noninteracting or interacting. Firstly, the DDI-2013 corpus was divided into a training set and a test set. For the sentence containing drugs d_i and d_j in the training set, three kinds of information [i.e. the word itself, part of speech (POS), relative distances between the word and each candidate drug in the sentence] of each word were extracted through the input layer, which were encoded into real-valued vectors (i.e. word embedding vectors, POS embedding vectors and position embedding vectors) by the embedding layer through looking up the corresponding embedding dictionary, respectively. Secondly, given that attention mechanisms could be used to quantify the effect of each word on the meaning of the sentence, the input attention layer was used to weigh the word embedding vectors. Thirdly, in the merging layer, each word was represented by a vector obtained by integrating the corresponding three types of embedding vectors. Then, the vectors representing the words were integrated into a sequence of vectors, which was input into the BLSTM layer to learn the global semantic representation of the sentence. Finally, the global representation of sentences was fed into the softmax layer to predict potential DDIs.

Position-aware multi-task deep learning method based on BLSTM

To automatically extract DDIs from biomedical texts, Zhou et al. [179] developed a position-aware multi-task deep learning method based on BLSTM (PM-BLSTM), whose architecture mainly consisted of four parts: embedding layer, BLSTM layer, position-aware attention layer and multi-task output layer. The DDI-2013 corpus was also used in this study, but unlike in Att-BLSTM where drug pairs were extracted directly from the sentences, for sentences involving more than two drugs, the authors used the rules proposed by Liu et al. [180] to filter the drugs to ensure that only one drug pair in each sentence was studied. For the sentence containing drugs d_i and d_j , through a word embedding dictionary and a position embedding dictionary, the embedding layer generated the corresponding embedding vector for each word. Then, the embedding vectors of all words were fed into the BLSTM layer to obtain the matrix composed of the hidden representation of each word. Considering that it was inaccurate to generate attentions only based on the local semantic information for a long sentence, the author utilized the position-aware attention layer to fuse the hidden representations of all words to get the sentence representation. Finally, the sentence representation was fed into the softmax-based classifiers in the multi-task output layer to identify potential DDIs.

A two-stage DDIs extraction model

Huang et al. [181] developed a two-stage method based on SVM and long short-term memory (LSTM) [182] to extract DDIs, where the SVM classifier was applied to identify potential DDIs, while the LSTM-based classifier was used for predicting the type of DDIs, including advise (two drugs are suggested to be taken together in the text), effect (the effects of two drugs taken together are described in the text), mechanism (the pharmacokinetic mechanism of DDI is introduced in the text) and int (there is no additional information about DDI in the text) according to the text description (Figure 8). Specifically, in the first stage, a feature definition approach [183] was used to extract the features for each sentence in the DDI-2013 corpus, including context word feature, pattern feature, verb feature, syntactic feature and auxiliary features. Then, the binary classifier SVM was used to classify the sentences into positive and negative instances. The drug pairs involved in the two instances were regarded as positive DDIs and

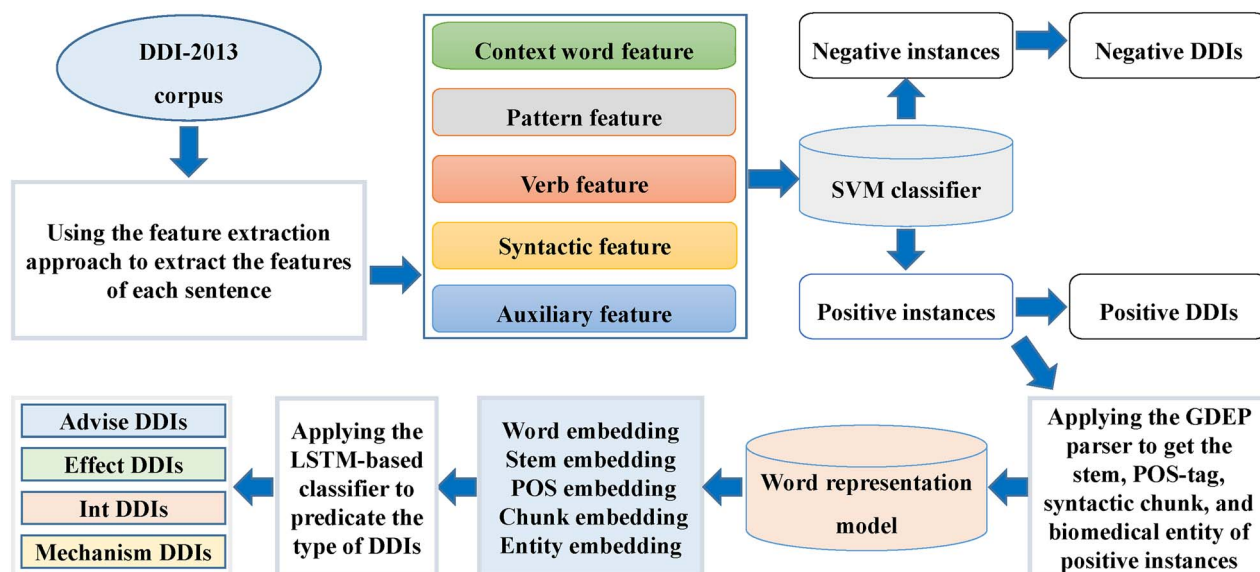


Figure 8. The flowchart of the two-stage DDIs extraction model built based on SVM and LSTM.

negative DDIs, respectively. In the second stage, the authors used the GDEP [184] parser to get the stem, POS-tag, syntactic chunk and biomedical entity of positive instances, based on which the word representation model [185] was applied to obtain the word embedding, stem embedding, POS embedding, chunk embedding and entity embedding of each word. Finally, the above embeddings of all words in each positive instance were fed into the LSTM-based classifier to predict the DDI type.

Instance position embedding and key external text for DDI extraction

Dou et al. [186] developed a framework named instance position embedding and key external text for DDI extraction (IK-DDI), where the instance position embedding was applied to extract DDI information from the DDI Extraction 2013 [187] database, while the key external text describing drugs was derived from the DrugBank. Specifically, firstly, the sentence (containing drugs d_i and d_j) recorded in DDI Extraction 2013 was input into the module (composed of the layers of Embedding, BiLSTM, CNN and MaxPooling) to obtain the feature vector f_{ij}^{int} of drug pair (d_i, d_j) . Secondly, given that the same drug may have different names in different texts, for drug d_i in the drug pair (d_i, d_j) , the authors first calculated the word string similarity SR_{ik} and word sense similarity SE_{ik} between the string of d_i and string of drug d_k in DrugBank according to the method presented in the study [188]. Then, the comprehensive similarity was defined:

$$S_{ik} = 2 \times \frac{SR_{ik} \times SE_{ik}}{SR_{ik} + SE_{ik}} \quad (66)$$

After calculating the comprehensive similarities between d_i and all drugs in DrugBank, the drug with the highest comprehensive similarity with d_i was regarded as the matched drug of d_i . The matched drug of d_j was obtained in a similar way. Next, the authors performed ‘Search key external text’ in DrugBank to mine two key sentences containing the matched drugs of d_i and d_j , respectively, which were input into the module consisting of the layers of Embedding, CNN and MaxPooling to get the feature vector f_{ij}^{ext} of drug pair (d_i, d_j) . Then, f_{ij}^{int} and f_{ij}^{ext} were input into the fully connected layer to obtain the final feature vector f_{ij} of the drug pair. Finally, based on f_{ij} , the softmax classification function

was applied to calculate the interaction probability between d_i and d_j .

3D graph and text-based neural network for drug–drug interaction prediction

By integrating the 3D GNN and pretrained text attention mechanism, Chen et al. [189] constructed a model named 3D graph and text-based neural network for drug–drug interaction prediction (3DGT-DDI). Firstly, through a force field optimization algorithm called MMFF [190], the authors obtained the 3D structure conformation of drug d_i based on corresponding SMILES, which was fed into the 3D GNN to get the structure-based feature of d_i . Then, the structure-based features of drug d_i and d_j were integrated as the feature vector of drug pair (d_i, d_j) . Secondly, sciBERT, a variant of bidirectional encoder representations from transformers (BERT) pretrained on scientific articles, was applied to tokenize the text describing drug pair (d_i, d_j) recorded in DDI Extraction 2013. Then, the tokenized text was used as the input of CNN to obtain the text-based feature vector of the drug pair. Finally, the above two types of feature vectors of the drug pair were input into the DNN to obtain the interaction probability between d_i and d_j .

Score function-based models

Given the successful application of score function-based models in the field of bioinformatics [42–44, 191], some researchers have developed models based on score functions to identify potential DDIs. The advantages of score function-based models are that the algorithm theory and calculation process involved are relatively easy to understand. Moreover, this type of model does not require negative samples. However, most of the score function-based models make predictions based on known DDIs, so they are not applicable to new drugs. In addition, when using this kind of model to predict DDIs, it is usually necessary to make assumptions about the probability distribution of DDIs, but if the data are inconsistent with the assumptions, the prediction accuracy of the models would be severely affected.

Russell–Rao-based model

To predict potential DDIs, Ferdousi et al. [192] proposed a computational model based on the Russell–Rao method [193]. To

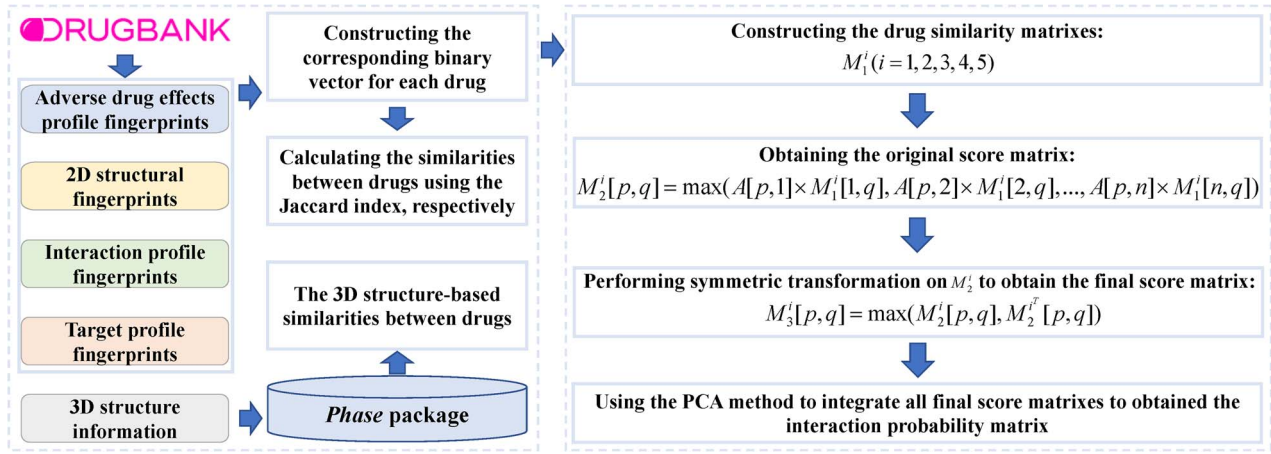


Figure 9. The flowchart of the score matrix and PCA-based model, where the score matrixes are calculated separately based on different similarities, and PCA is used to integrate all the score matrices into the final interaction probability matrix.

be specific, according to the known associations between drugs and four types of biological elements (including 23 carriers, 115 transporters, 235 enzymes and 1787 targets), the authors constructed four corresponding binary vectors for each drug. Given that there were shared proteins between these four types of biological elements, the authors spliced four types of binary vectors to construct the comprehensive binary vector (2004 dimension) for each drug after removing the redundant proteins. Then, the method of Russell–Rao [193] was used to calculate the interaction probability $P(d_i, d_j)$ between drugs d_i and d_j :

$$P(d_i, d_j) = \frac{V_{d_i}^T V_{d_j}}{2004} \quad (67)$$

where V_{d_i} and V_{d_j} represent the comprehensive binary vectors of drug d_i and d_j , respectively.

Score matrix and PCA-based model

Vilar et al. [194] proposed a DDIs prediction model by constructing score matrixes according to adjacency matrix and similarity matrixes (Figure 9). Firstly, based on the drug-related information (including 2D structural fingerprints [195], interaction profile fingerprints [196], target profile fingerprints [55] and adverse drug effects (ADEs) profile fingerprints [197]) downloaded from DrugBank, the authors constructed the corresponding binary vector for each drug and then calculated the similarities between drugs using the Jaccard index, respectively. Besides, the authors calculated the 3D structure-based similarities between drugs by the Phase package. Then, five corresponding similarity matrixes were constructed, which were represented by M_1^i ($i = 1, 2, 3, 4, 5$). Secondly, they defined the original score matrix M_2^i as follows:

$$M_2^i[p, q] = \max(A[p, 1] \times M_1^i[1, q], A[p, 2] \times M_1^i[2, q], \dots, A[p, n] \times M_1^i[n, q]) \quad (68)$$

where n refers to the total number of drugs and A represents the adjacency matrix constructed based on known DDIs. Since the matrix M_2^i is not symmetric, the authors performed symmetric transformation on the matrix M_2^i to obtain the final score matrix M_3^i :

$$M_3^i[p, q] = \max(M_2^i[p, q], M_2^{i^T}[p, q]) \quad (69)$$

Finally, the PCA method was used to integrate the score matrixes (i.e. $M_3^1, M_3^2, M_3^3, M_3^4, M_3^5$) to obtain the interaction probability matrix.

DISCUSSION AND CONCLUSION

In clinical treatment, in order to cure the disease as soon as possible, patients usually take two or more drugs. The combination of some drugs could not only increase the efficacy of drugs but also delay the emergence of resistance [198]. However, inappropriate drug combinations not only fail to achieve the expected therapeutic effect but also may cause adverse reactions, even toxic reactions. Many drugs are forced to stop selling due to serious adverse reactions caused by DDIs, which not only brings harm to patients but also brings huge economic losses to pharmaceutical companies. More than 20 years ago, the calcium channel blocker mibefradil was withdrawn from the market because it could lead to lethal DDIs by inhibiting the cytochrome P450 3A4 metabolism of certain drugs [199]. For this reason, manufacturers are required to specify DDIs strictly in drug instructions, and consumers must carefully read the instructions when using drugs. As more and more new drugs are approved for clinical treatment, the number of potential DDIs increases rapidly. However, due to the time and money constraints, a large number of potential DDIs that may cause adverse reactions are not provided in drug instructions. With the deepening of the understanding of drug metabolism mechanisms and the rapid accumulation of drug-related data, more and more researchers were committed to building computational models to predict potential DDIs.

In this review, we first introduced the basic conception and classification of DDIs. Some important publicly available databases and web servers about experimentally verified or predicted DDIs were also briefly described. Besides, we summarized three types of prediction models proposed during recent years and discussed the advantages as well as limitations of them. Finally, we pointed out the problems that need to be solved in the future research of DDIs prediction and provided corresponding suggestions. In general, this review is helpful for researchers to have a comprehensive understanding of DDIs prediction and provides valuable guidance for their research studies, especially in the construction of models; they can weigh the advantages and disadvantages of various models to build the most suitable model for their own research studies.

Several researchers have written reviews to summarize the DDIs prediction models. For example, Zhang *et al.* [200] summarized deep learning-based models for extracting DDIs from the literatures and divided them into three categories: CNN-based model, RNN-based model and recursive NN-based model. Then, the authors compared the performance of these models in the DDI corpus and summarized the strengths as well as weaknesses of these models. Finally, the authors discussed the challenges and future prospects of extracting DDIs by deep learning-based models. Besides, Lin *et al.* [201] first listed the DDIs prediction models constructed based on deep learning as well as graph learning and evaluated their performance based on different tasks, including the binary classification task, multi-class classification task as well as multi-label classification task. In addition, they introduced a variety of molecular representation methods of drugs, such as sequence-based, 2D graph-based, 3D graph-based, knowledge graph-based and so on. Finally, the authors discussed the potential technical challenges and highlighted the future directions of predicting DDIs by the above two types of models.

Although both reviews provided detailed summaries of models for DDIs prediction, they mainly focused on deep learning-based models. There were other types of models that were built to predict potential DDIs, such as traditional machine learning-based models and score function-based models. In order to enable researchers to have a more comprehensive understanding of the research studies related to DDIs prediction, we summarized the above three types of models in this review.

There were obvious differences among the above three types of models. For example, the traditional machine learning-based models were designed to leverage classical machine learning algorithms to make reliable predictions by building efficient features or solving specific optimization problems. The deep learning-based models were designed to automatically learn the significant features of drug pairs and then implement DDIs prediction. In the score function-based models, the authors constructed the corresponding function based on probability distribution or statistical analysis to calculate the interaction probabilities between drugs. Besides, different from the other two types of models used to implement DDIs prediction based only on the data about drug similarity and known DDIs, most deep learning-based models also utilized substructure information about drugs, and the models built based on NLP methods were applied to mine potential DDIs based on the texts in the literatures. Moreover, in the deep learning-based models, some specified modules (including self-attention mechanism, Siamese network and so on) were used to fuse the features of two drugs to obtain the features of the corresponding drug pair, rather than splicing the features of two drugs directly in the other two types of models. Compared with the other two types of models, the score function-based models were easier to understand and does not require negative samples. We briefly summarized each type of model below.

A variety of classical algorithms were involved in the traditional machine learning-based models, such as label propagation, MRFs, random forest, logistic regression, SVM, matrix factorization, ensemble learning and so on. For example, in INDI [57], the authors first calculated seven types of drug similarities and obtained the scores between two drugs through a scoring scheme based on the drug similarities. Then, according to the scores between two drugs, the feature vector of corresponding drug pair was constructed, which was used as the input of logistic regression classifier to implement DDIs predictions. In the DDIs prediction model built by Liu *et al.* [73] based on random forest, the authors first constructed the feature vector for each drug

pair according to the information about the chemical structure and target protein of drugs. Then, the feature vector was fed into the random forest model to calculate the interaction probability between the corresponding drugs.

In the deep learning-based model, researchers applied different techniques (i.e. CNNs, DNNs, GNNs, Graph Embedding and so on) to predict DDIs based on drug-related information or used relevant methods of NLP to mine potential DDIs from texts. Taking MCFF-MTDDI [146] as an example, after obtaining the chemical structure-based feature and KG fusion representations of each drug pairs, the two types of features were input into a GRU-based multichannel feature fusion framework to obtain the fused feature vector, which was used as the input of the multi-class classifier to implement the multi-class classification tasks. Besides, the extra label-based feature vector and fused feature vector of drug pair were concatenated and input into the multi-label classifier to implement the multi-label classification. In the model constructed by Huang *et al.* [181] based on the two-stage method, after defining the features of each sentence, the authors first divided the sentences into positive instances and negative instances using SVM. Then, they used the word representation model to obtain multiple embedding of each word in the positive instances, which was fed into the LSTM-based classifier to predict the type of the corresponding DDIs.

In the score function-based models, the authors defined score functions from different perspectives to calculate the interaction probabilities between drugs. For example, in the model constructed based on the Russell-Rao method, according to four kinds of association information about drugs, the authors constructed corresponding binary vectors for each drug, which were integrated into a comprehensive binary vector. Finally, the Russell-Rao method was used to calculate the interaction probability based on the comprehensive binary vectors of two drugs.

Next, we summarized the advantages and disadvantages of the above three types of models. The traditional machine learning-based model can be used to make large-scale rapid prediction for potential DDIs. Besides, the main advantage of this type of model is that they are suitable for new drugs, such as TMFUF [99], which can be used to predict DDIs between new drugs. However, they still have some limitations. For example, the traditional machine learning-based models involve multiple parameters, and the setting of parameter values limits the model performance to some extent. Besides, researchers tended to define the feature vectors of drug pairs based on the similarities between drug, so constructing feature vectors with higher significance is still an urgent problem to be solved. Moreover, given that negative samples were extremely difficult to obtain, researchers treated candidate samples as negative samples to train this type of models, which limited the performance of the model to some extent. Compared with traditional machine learning methods, deep learning methods can automatically mine the significant features of drugs. In addition, deep learning-based models have high flexibility in feature fusion. For instance, in MDF-SA-DDI [159], the self-attention mechanism was used to fuse the feature vectors of drug pairs, while a GRU-based multichannel feature fusion framework was applied to integrate the feature vectors of drug pairs in MCFF-MTDDI [146]. Compared with the simple splicing of feature vectors, the above fusion methods can fully fuse features. However, implementing predictions with deep learning-based models often takes more time. As with models constructed based on the traditional machine learning algorithms, the scarcity of reliable negative samples severely limits the performance of models. Besides, deep learning-based models

lack interpretability. The advantage of score function-based models is that the algorithm theory and calculation process involved are relatively easy to understand. Moreover, this type of model does not require negative samples. However, they still have some deficiencies. For instance, most of the score function-based models are not suitable for new drugs. In addition, when these models are applied to predict potential DDIs, assumptions about the probability distribution are often required, but if the data are inconsistent with the assumptions, the prediction accuracy of the models will be severely affected.

Considering the advantages and disadvantages of each type of models, in my opinion, it is best for researchers to build models based on deep learning algorithms in the future research. Compared with the other two types of models, the computational efficiency of deep learning-based models is lower, but their accuracy is generally higher. Besides, deep learning-based models can predict interactions between drug substructures, which is beneficial for us to understand the mechanism of DDIs. NLP models could not only be used to mine a large number of DDIs from the literatures but also help researchers learn detailed information about DDIs. As for the disadvantages of this type of model, instead of randomly selecting drug pairs without known interactions as negative samples, designing methods to identify reliable negative samples is more conducive to further improving the accuracy of the models. Given the lack of interpretability in most deep learning-based models, it is necessary to perform interpretability analysis on these models. In addition, in view of the fact that the values of parameters in models affect the prediction performance to a certain extent, designing the model to determine the optimal values of the parameters helps further improve the prediction accuracy.

To evaluate the predictive performance of computational models, most researchers conducted k-fold cross validation and case studies. Besides, there are other methods used to evaluate performance. For example, in MCFF-MTDDI [146], Chen *et al.* applied seven classical evaluation indicators (including Accuracy, Macro-Precision, Macro-Recall, Macro-F1, Cohen's Kappa, AUC and AUPR) to evaluate the prediction performance of model in a multi-class classification task, while AUC and AUPR were used to evaluate the performance of model in multi-label prediction tasks. In BioDKG-DDI [154], the Matthews's correlation coefficient (MCC) was used to evaluate the effectiveness of the model. These indicators are very convincing in assessing the performance of DDIs prediction models. Besides, some researchers performed ablation studies to assess the effect of each module on predictive performance. For example, in SSI-DDI [129], the authors removed the co-attention layer and changed the number of GAT layers to evaluate the effect of the corresponding two modules on the prediction accuracy, respectively.

In view of that discovering potential DDIs would be beneficial to drug development and clinical treatment, researchers have developed several DDI prediction models with superior performance. However, there are still some problems to be solved in the future. Firstly, the data are extremely unbalanced, i.e. the number of positive samples is much smaller than the number of candidate samples. Therefore, it is necessary to collect more drug-drug pairs with known interactions as positive samples in the future work. Secondly, current research studies focus on the identification of potential DDIs and the prediction of DDIs types, but less research has been done on the severity of DDIs. Besides, the existing prediction models do not take into account the effect of drug dose on DDIs. Thirdly, current models can only be used to predict interactions between two drugs, which is far from

enough. In clinical treatment, patients often need to take more than two drugs. Therefore, it is of great significance to study the interactions among multiple drugs. In order to solve the above two problems, it is necessary to build the corresponding databases to provide the data foundation for the subsequent research studies. Fourthly, as more and more drug-related data are generated, applying deep learning algorithms to effectively merge them is expected to further improve the prediction accuracy. Finally, many models made predictions only based on the similarity and association information of drugs, but these models could not reveal the mechanism of interactions. If text information describing the drugs is introduced, this problem is expected to be improved.

Key Points

- We introduced the basic conception and classification of DDIs. In addition, several databases and web servers about DDIs were introduced.
- Paying attention to DDIs is of great significance to adopt effective combination therapy and further improve the quality of medical treatment.
- Revealing DDIs through experiments is extremely time-consuming and costly, and building the computational models to calculate the interaction probabilities between drugs could be an important complement to experimental methods.
- Based on the calculation principle of the models, we simply divided the models into three categories: traditional machine learning-based models, deep learning-based models and score function-based models.
- We briefly discussed the advantages and limitations of existing computational models and put forward the existing problems in the current DDIs prediction research, which need to be resolved in the future work.

FUNDING

This work was funded by National Natural Science Foundation of China under Grant No. 92370131, 61972399, 11931008 and the Postgraduate Research & Practice Innovation Program of Jiangsu Province and the Postgraduate Research & Practice Innovation Program of China University of Mining and Technology (No. KYCX19_2180).

DATA AVAILABILITY

The source code and data of Bayesian probabilistic method-based model are available at <http://www.picb.ac.cn/hanlab/DDI>. The source code and data of LCM-DS are available at <https://github.com/JustinShi2016/ScientificReports2018>. The source code and data of Network algorithm and matrix perturbation algorithm-based model are available at <https://github.com/zw9977129/drug-drug-interaction/>. The source code and data of SFLN are available at <https://github.com/BioMedicalBigDataMiningLabWhu/SFLN>. The source code and data of DDIMDL are available at <https://github.com/YifanDengWHU/DDIMDL>. The source code and data of SSI-DDI are available at <https://github.com/kanz76/SSI-DDI>. The source code and data of STNN-DDI are available at <https://github.com/zsy-9/STNN-DDI>. The source code and data of META-DDIE are available at <https://github.com/YifanDengWHU/META-DDIE>. The source code and data of DANN-DDI are available at <https://github.com/naodandandan/DANN-DDI>. The source

code and data of MRCGNN are available at <https://github.com/Zhankun-Xiong/MRCGNN>. The source code and data of MCFF-MTDDI are available at <https://github.com/ChendiHan111/MCFF-MTDDI>. The source code and data of DSN-DDI are available at <https://github.com/microsoft/Drug-Interaction-Research/tree/DSN-DDI> for-DDI-Prediction. The source code and data of MDF-SA-DDI are available at <https://github.com/ShenggengLin/MDF-SA-DDI>. The source code and data of R2-DDI are available at <https://github.com/linjc16/R2-DDI>. The source code and data of Graph kernel-based approach are available at <https://sbmi.uth.edu/ccb/resources/ddi.htm>. The source code and data of IK-DDI are available at <https://github.com/DouMingLiang/IK-DDI>. The source code and data of 3DGT-DDI are available at <https://github.com/hehh77/3DGT-DDI>.

REFERENCES

1. Dale MM, Haylett DG. *Rang & Dale's Pharmacology Flash Cards Updated Edition E-Book*. Elsevier Health Sciences, London, UK, 2013.
2. Younger P. *Stedman's Medical Dictionary*. Ref Rev. Lippincott Williams & Wilkins, Baltimore, USA, 2007.
3. Dictionary AH. *The American Heritage Science Dictionary*. Houghton Mifflin Harcourt, Boston, USA, 2005.
4. Pereira RS. The use of baker's yeast in the generation of asymmetric centers to produce chiral drugs and others compounds. *Crit Rev Biotechnol* 1998;**18**:25–64.
5. Atanasov AG, Waltenberger B, Pferschy-Wenzig E-M, et al. Discovery and resupply of pharmacologically active plant-derived natural products: a review. 2015;**33**:1582–614.
6. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
7. Zheng S, Aldahdooh J, Shadbahr T, et al. DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;**49**:W174–84.
8. Espinal MA, Kim SJ, Suarez PG, et al. Standard short-course chemotherapy for drug-resistant tuberculosis: treatment outcomes in 6 countries. *JAMA* 2000;**283**:2537–45.
9. Walkup JT, Albano AM, Piacentini J, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med* 2008;**359**:2753–66.
10. Keith CT, Boris AA, Stockwell BR. Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov* 2005;**4**:71–8.
11. Keith CT, Boris AA, Stockwell BR. Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov* 2005;**4**:71–8.
12. Genina N, Boetker JP, Colombo S, et al. Anti-tuberculosis drug combination for controlled oral delivery using 3D printed compartmental dosage forms: from drug product design to in vivo testing. *J Control Release* 2017;**268**:40–8.
13. Huang J, Niu C, Green CD, et al. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Comput Biol* 2013;**9**:e1002998.
14. Qato DM, Wilder J, Schumm LP, et al. Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the United States, 2005 vs 2011. *JAMA Intern Med* 2016;**176**:473–82.
15. Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov* 2005;**4**:825–33.
16. Juurlink DN, Mamdani M, Kopp A, et al. Drug-drug interactions among elderly patients hospitalized for drug toxicity. *JAMA* 2003;**289**:1652–8.
17. Yeh P, Tschumi AI, Kishony R. Functional classification of drugs by properties of their pairwise interactions. *Nat Genet* 2006;**38**:489–94.
18. Zhao XM, Iskar M, Zeller G, et al. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 2011;**7**:e1002323.
19. Beijnen JH, Schellens JH. Drug interactions in oncology. *Lancet Oncol* 2004;**5**:489–96.
20. Papaseit E, Pérez-Mañá C, Torrens M, et al. MDMA interactions with pharmaceuticals and drugs of abuse. *Expert Opin Drug Metab Toxicol* 2020;**16**:357–69.
21. Finerman GA, Milch RA. In vitro binding of tetracyclines to calcium. *Nature* 1963;**198**:486–7.
22. Kantrowitz PA, Siegel CI, Strong MJ, et al. Response of the human oesophagus to d-tubocurarine and atropine. *Gut* 1970;**11**:47–50.
23. Scripture CD, Figg WD. Drug interactions in cancer therapy. *Nat Rev Cancer* 2006;**6**:546–58.
24. Ray WA, Chung CP, Murray KT, et al. Association of proton pump inhibitors with reduced risk of warfarin-related serious upper gastrointestinal bleeding. *Gastroenterology* 2016;**151**:1105–1112.e10.
25. Leape LL, Bates DW, Cullen DJ, et al. Systems analysis of adverse drug events, ADE prevention study group. *JAMA* 1995;**274**:35–43.
26. Day RO, Snowden L, McLachlan AJ. Life-threatening drug interactions: what the physician needs to know. *Intern Med J* 2017;**47**:501–12.
27. Preskorn SH. Drug-drug interactions in psychiatric practice, part 1: reasons, importance, and strategies to avoid and recognize them. *J Psychiatr Pract* 2018;**24**:261–8.
28. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today* 2005;**10**:1421–33.
29. Xiong G, Yang Z, Yi J, et al. DDInter: an online drug-drug interaction database towards improving clinical decision-making and patient safety. *Nucleic Acids Res* 2022;**50**:D1200–d1207.
30. Siramshetty VB, Eckert OA, Gohlke BO, et al. SuperDRUG2: a one stop resource for approved/marketed drugs. *Nucleic Acids Res* 2018;**46**:D1137–d1143.
31. Bottiger Y, Laine K, Andersson ML, et al. SFINX-a drug-drug interaction database designed for clinical decision support systems. *Eur J Clin Pharmacol* 2009;**65**:627–33.
32. Yap KY, Kuo EY, Lee JJ, et al. An onco-informatics database for anticancer drug interactions with complementary and alternative medicines used in cancer treatment and supportive care: an overview of the OncoRx project. *Support Care Cancer* 2010;**18**:883–91.
33. Hachad H, Ragueneau-Majlessi I, Levy RH. A useful tool for drug interaction evaluation: the University of Washington Metabolism and Transport Drug Interaction Database. *Hum Genomics* 2010;**5**:61–72.
34. Liu Y, Liang Y, Wishart D. PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res* 2015;**43**:W535–42.
35. Luo H, Zhang P, Huang H, et al. DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome. *Nucleic Acids Res* 2014;**42**:W46–52.
36. Schyman P, Liu R, Desai V, Wallqvist A. vNN web server for ADMET predictions. *Front Pharmacol* 2017;**8**:889.
37. Cheng D, Knox C, Young N, et al. PolySearch: a web-based text mining system for extracting relationships between human

- diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Res* 2008;**36**:W399–405.
38. Yang L, Chen J, Shi L, et al. Identifying unexpected therapeutic targets via chemical-protein interactome. *PloS One* 2010;**5**:e9568.
 39. Liu R, Tawa G, Wallqvist A. Locally weighted learning methods for predicting dose-dependent toxicity with application to the human maximum recommended daily dose. *Chem Res Toxicol* 2012;**25**:2216–26.
 40. Guh RS, Shiue YRJ. An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts. *Computers & Industrial Engineering* 2008;**55**: 475–93.
 41. Yan W, Shao H, Wang X. Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers & Chemical Engineering* 2004;**28**:1489–98.
 42. Chen X, Yan CC, Zhang X, You ZH. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2017;**18**:558–76.
 43. Chen X, Xie D, Zhao Q, You ZH. MicroRNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2019;**20**:515–39.
 44. Chen X, Guan NN, Sun YZ, et al. MicroRNA-small molecule association identification: from experimental results to computational models. *Brief Bioinform* 2020;**21**:47–61.
 45. Wang CC, Zhao Y, Chen X. Drug-pathway association prediction: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbaa061.
 46. Chen X, Sun YZ, Liu H, et al. RNA methylation and diseases: experimental results, databases, web servers and computational models. *Brief Bioinform* 2019;**20**:896–917.
 47. Chen X, Yan CC, Zhang X, et al. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**:696–712.
 48. Huang L, Zhang L, Chen X. Updated review of advances in microRNAs and complex diseases: taxonomy, trends and challenges of computational models. *Brief Bioinform* 2022;**23**:bbac358.
 49. Huang L, Zhang L, Chen X. Updated review of advances in microRNAs and complex diseases: towards systematic evaluation of computational models. *Brief Bioinform* 2022;**23**:bbac407.
 50. Huang L, Zhang L, Chen X. Updated review of advances in microRNAs and complex diseases: experimental results, databases, web servers and data fusion. *Brief Bioinform* 2022;**23**:bbac397.
 51. Wang CC, Han CD, Zhao Q, Chen X. Circular RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbab286.
 52. Zhao Y, Wang CC, Chen X. Microbes and complex diseases: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbaa158.
 53. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human protein reference database–2009 update. *Nucleic Acids Res* 2009;**37**:D767–72.
 54. Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 2004;**101**:6062–7.
 55. Campillos M, Kuhn M, Gavin AC, et al. Drug target identification using side-effect similarity. *Science* 2008;**321**:263–6.
 56. Xia K, Dong D, Han JD. IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model. *BMC Bioinformatics* 2006;**7**:508.
 57. Gottlieb A, Stein GY, Oron Y, et al. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 2012;**8**:592.
 58. Steinbeck C, Hoppe C, Kuhn S, et al. Recent developments of the chemistry development kit (CDK)-an open-source java library for chemo- and bioinformatics. *Curr Pharm Des* 2006;**12**: 2111–20.
 59. Keiser MJ, Roth BL, Armbruster BN, et al. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**:197–206.
 60. Atias N, Sharan R. An algorithmic framework for predicting side effects of drugs. *J Comput Biol* 2011;**18**:207–18.
 61. 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.
 62. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 2009;**25**:2397–403.
 63. Perlman L, Gottlieb A, Atias N, et al. Combining drug and gene similarity measures for drug-target elucidation. *J Comput Biol* 2011;**18**:133–45.
 64. 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**:496.
 65. Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**:12339.
 66. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016;**44**:D1075–9.
 67. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**:125ra131.
 68. Wang Y, Xiao J, Suzek TO, et al. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009;**37**:W623–33.
 69. Wang F, Li P, König AC, Wan M. Improving clustering by learning a bi-stochastic data similarity matrix. *Knowledge and Information Systems* 2012;**32**:351–82.
 70. Ciuciu P, Idier J. A half-quadratic block-coordinate descent method for spectral estimation. *Signal Processing* 2002;**82**: 941–59.
 71. Sridhar D, Fakhraei S, Getoor L. A probabilistic approach for collective similarity-based drug-drug interaction prediction. *Bioinformatics* 2016;**32**:3175–82.
 72. Bach SH. *Hinge-Loss Markov Random Fields and Probabilistic Soft Logic: A Scalable Approach to Structured Prediction*. College Park: University of Maryland, 2015.
 73. Liu L, Chen L, Zhang YH, et al. Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection. *J Biomol Struct Dyn* 2017;**35**:312–29.
 74. Chen L, Lu J, Zhang N, et al. A hybrid method for prediction and repositioning of drug anatomical therapeutic chemical classes. *Mol Biosyst* 2014;**10**:868–77.
 75. Szklarczyk D, Santos A, von Mering C, et al. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 2016;**44**:D380–4.
 76. Jensen LJ, Kuhn M, Stark M, et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009;**37**:D412–6.
 77. Huang T, Zhang J, Xu ZP, et al. Deciphering the effects of gene deletion on yeast longevity using network and machine learning approaches. *Biochimie* 2012;**94**:1017–25.

78. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;**27**: 1226–38.
79. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 1975;**405**:442–51.
80. Witten IH, Frank E, Hall MA. *Data Mining : Practical Machine Learning Tools and Techniques*, Third Edition. ACM SIGSOFT Software Engineering Notes 2011;**5**:51–2.
81. Takeda T, Hao M, Cheng T, et al. Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. *J Chem* 2017;**9**:16.
82. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res* 2015;**44**:D1202–13.
83. Kuhn M. *Caret: Classification and Regression Training*. Astrophysics Source Code Library, 2015.
84. Hameed PN, Verspoor K, Kusljic S, Halgamuge S. Positive-unlabeled learning for inferring drug interactions based on heterogeneous attributes. *BMC Bioinformatics* 2017;**18**:140.
85. Li J, Lu Z. A new method for computational drug repositioning using drug pairwise similarity. In: *2012 IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, USA, 2012, p. 1–4. IEEE, New York, NY, USA.
86. Kuhn M, Campillos M, Letunic I, et al. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;**6**:343.
87. Chan C-KK, Hsu AL, Halgamuge SK, Tang SL. Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics* 2008;**9**:215.
88. Deepika SS, Geetha TV. A meta-learning framework using representation learning to predict drug-drug interaction. *J Biomed Inform* 2018;**84**:136–47.
89. Zhuang F, Zhang Z, Qian M, et al. Representation learning via dual-autoencoder for recommendation. *Neural Netw* 2017;**90**: 83–9.
90. Yang P, Li XL, Mei JP, et al. Positive-unlabeled learning for disease gene identification. *Bioinformatics* 2012;**28**:2640–7.
91. Lemke C, Budka M, Gabrys B. Metalearning: a survey of trends and technologies. *Artif Intelligence Rev* 2015;**44**:117–30.
92. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, p. 855–64. Association for Computing Machinery, New York, NY, USA.
93. Zhang W, Chen Y, Li D, Yue X. Manifold regularized matrix factorization for drug-drug interaction prediction. *J Biomed Inform* 2018;**88**:90–7.
94. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* 2000;**290**:2323–6.
95. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;**290**:2319–23.
96. Yu H, Mao KT, Shi JY, et al. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol* 2018;**12**:14.
97. Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In: *Proceedings of the 13th International Conference on Neural Information Processing Systems*, Denver, USA, 2000, p. 535–41. MIT Press, Cambridge, UK.
98. De Jong S. SIMPLS: an alternative approach to partial least squares regression. *Chemom Intel Lab Syst* 1993;**18**:251–63.
99. Shi JY, Huang H, Li JX, et al. TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs. *BMC Bioinformatics* 2018;**19**:411.
100. Shi JY, Shang XQ, Gao K, et al. An integrated local classification model of predicting drug-drug interactions via Dempster-Shafer theory of evidence. *Sci Rep* 2018;**8**:11829.
101. Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**:12339.
102. Suykens J, Vandewalle J. Least squares support vector machine classifiers. 1999;**9**:293–300.
103. Mesarovic VZ, Galatsanos NP, Katsaggelos AK. Regularized constrained total least squares image restoration. *IEEE Transactions on Image Processing* 1995;**4**:1096–108.
104. Lucena D, Prudencio R. Semi-supervised multi-label k-nearest neighbors classification algorithms. In: *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, Natal, Brazil, 2016, p. 49–54. IEEE, New York, NY, USA.
105. Beynon M, Curry B, Morgan PJO. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega* 2000;**28**:37–50.
106. Yan C, Duan G, Pan Y, et al. DDIGIP: predicting drug-drug interactions based on Gaussian interaction profile kernels. *BMC Bioinformatics* 2019;**20**:538.
107. Qian S, Liang S, Yu H. Leveraging genetic interactions for adverse drug-drug interaction prediction. *PLoS Comput Biol* 2019;**15**:e1007068.
108. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;**32**:267D–0.
109. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf* 1999;**20**:109–17.
110. Zhao M, Lee WP, Garrison EP, Marth GT. SSW library: an SIMD Smith-Waterman C/C++ library for use in genomic applications. *PLoS One* 2013;**8**:e82138.
111. Costanzo M, VanderSluis B, Koch EN, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* 2016;**353**:aaf1420.
112. Breheny P, Huang J. Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Stat Comput* 2015;**25**:173–87.
113. Bergstra J, Bardenet R, Bengio Y et al. Algorithms for hyperparameter optimization. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, 2011, p. 2546–54. NIPS, New York, NY, USA.
114. Zhang W, Chen Y, Liu F, et al. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics* 2017;**18**:18.
115. Bobadilla J, Ortega F, Hernando A, Gutiérrez A. Recommender systems survey. *Knowl-Based Syst* 2013;**46**:109–32.
116. Chen X, Liu MX, Yan GY. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol Biosyst* 2012;**8**: 1970–8.
117. Lü L, Pan L, Zhou T, et al. Toward link predictability of complex networks. *Proc Natl Acad Sci U S A* 2015;**112**:2325–30.
118. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J Am Med Inform Assoc* 2014;**21**:e278–86.
119. Cheng F, Li W, Wu Z, et al. Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space. *J Chem Inf Model* 2013;**53**:753–62.

120. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* 2006;**11**:1046–53.
121. Hunta S, Yooyativong T, Aunsri N. A novel integrated action crossing method for drug-drug interaction prediction in non-communicable diseases. *Comput Methods Programs Biomed* 2018;**163**:183–93.
122. Zhang W, Jing K, Huang F, et al. SFLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inform Sci* 2019;**497**: 189–201.
123. Zhang W, Chen Y, Tu S, et al. Drug side effect prediction through linear neighborhoods and multiple data source integration. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Shenzhen, China, 2016, p. 427–34. IEEE, New York, NY, USA.
124. Bouraoui A, Jamoussi S, Hamadou AB. A comprehensive review of deep learning for natural language processing. *Int J Data Min Model Manag* 2022;**14**:149–82.
125. Otter DW, Medina JR, Kalita JK. A survey of the usages of deep learning for natural language processing. *IEEE Trans Neural Netw Learn Syst* 2021;**32**:604–24.
126. Shahidi Zandi M, Rajabi R. Deep learning based framework for Iranian license plate detection and recognition. *Multimed Tools Appl* 2022;**81**:15841–58.
127. Xing S, Jiao Y, Salehzadeh M, et al. SteroidXtract: deep learning-based pattern recognition enables comprehensive and rapid extraction of steroid-like metabolic features for automated biology-driven metabolomics. *Anal Chem* 2021;**93**:5735–43.
128. Deng Y, Xu X, Qiu Y, et al. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics* 2020;**36**:4316–22.
129. Nyamabo AK, Yu H, Shi JY. SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction. *Brief Bioinform* 2021;**22**:bbab133.
130. Lee J, Lee I, Kang J. Self-attention graph pooling. In: *Proceedings of the 36th International Conference on Machine Learning*, California, USA, 2019, p. 3734–43. ACM, New York, NY, USA.
131. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *International Conference on Learning Representations*, San Diego, USA, 2014. ICLR, Washington DC, USA.
132. Yu H, Zhao S, Shi J. STNN-DDI: a substructure-aware tensor neural network to predict drug-drug interactions. *Brief Bioinform* 2022;**23**:bbac209.
133. Goulart J, Boizard M, Boyer R, et al. Tensor CP decomposition with structured factor matrices: algorithms and performance. *IEEE J Sel Top Signal Processing* 2016;**10**:757–69.
134. Chen H, Vorobyov SA, So HC, et al. Introduction to the special issue on tensor decomposition for signal processing and machine learning. *IEEE Journal of Selected Topics in Signal Processing* 2021;**15**:433–7.
135. Deng Y, Qiu Y, Xu X, et al. META-DDIE: predicting drug-drug interaction events with few-shot learning. *Brief Bioinform* 2022;**23**:bbab514.
136. Cellier P, Charnois T, Plantevit M, et al. Sequential pattern mining for discovering gene interactions and their contextual information from biomedical texts. *J Biomed Semantics* 2015;**6**:27.
137. Huang K, Xiao C, Hoang TN et al. CASTER: predicting drug interactions with chemical substructure representation. In: *AAAI Conference on Artificial Intelligence*. 2019, **34**, p. 702–9. AAAI Press, Palo Alto, USA.
138. Sung F, Yang Y, Zhang L et al. Learning to compare: relation network for few-shot learning. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018, p. 1199–208. IEEE, New York, NY, USA.
139. Liu S, Zhang Y, Cui Y, et al. Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform* 2023;**20**:976–85.
140. Liu S, Huang Z, Qiu Y, et al. Structural network embedding using multi-modal deep auto-encoders for predicting drug-drug. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, USA, 2019, p. 445–50. IEEE, New York, NY, USA.
141. Zhang P, Zhao B, Wong L et al. A novel computational method for predicting LncRNA-disease associations from heterogeneous information network with SDNE embedding model. In: *Intelligent Computing Theories and Application: 16th International Conference*, Bari, Italy, 2020, p. 505–13. Springer-Verlag, Heidelberg, Germany.
142. Qu M, Tang J, Shang J et al. An attention-based collaboration framework for multi-view network representation learning. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore, Singapore: Association for Computing Machinery, 2017, p. 1767–76. ACM, New York, NY, USA.
143. Xiong Z, Liu S, Huang F et al. Multi-Relational Contrastive Learning Graph Neural Network for Drug-Drug Interaction Event Prediction. In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, Washington DC, USA, 2023, p. 5339–47. Menlo Park, USA.
144. Li P, Li Y, Hsieh CY, et al. TrimNet: learning molecular representation from triplet messages for biomedicine. *Brief Bioinform* 2021;**22**:bbaa266.
145. Schlichtkrull M, Kipf T, Bloem P et al. Modeling relational data with graph convolutional networks. In: *Extended Semantic Web Conference*, Heraklion, Greece, 2018, p. 593–607. Springer, Cham, Switzerland.
146. Han CD, Wang CC, Huang L, Chen X. MCFF-MTDDI: multi-channel feature fusion for multi-typed drug-drug interaction prediction. *Brief Bioinform* 2023;**24**:bbad215.
147. Al-Saleem J, Granet R, Ramakrishnan S, et al. Knowledge graph-based approaches to drug repurposing for COVID-19. *J Chem Inf Model* 2021;**61**:4058–67.
148. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;**50**:742–54.
149. Tang Z, Chen G, Yang H, et al. DSIL-DDI: a domain-invariant substructure interaction learning for generalizable drug–drug interaction prediction. *IEEE Trans Neural Netw Learn Syst* 2023; 1–9.
150. Schölkopf B, Locatello F, Bauer S, et al. Toward causal representation learning. *Proceedings of the IEEE* 2021;**109**:612–34.
151. Li Z, Zhu S, Shao B, et al. DSN-DDI: an accurate and generalized framework for drug-drug interaction prediction by dual-view representation learning. *Brief Bioinform* 2023;**24**: bbac597.
152. Yang L, Li W, Guo Y, Gu J. Graph-CAT: graph co-attention networks via local and global attribute augmentations. *Future Generation Computer Systems* 2021;**118**:170–9.
153. Edelman BL, Goel S, Kakade S et al. Inductive biases and variable creation in self-attention mechanisms. In: *Proceedings of the 39th International Conference on Machine Learning*, Baltimore, USA, 2022, p. 5793–831. PMLR, New York, NY, USA.
154. Ren ZH, Yu CQ, Li LP, et al. BioDKG-DDI: predicting drug-drug interactions based on drug knowledge graph fusing biochemical information. *Brief Funct Genomics* 2022;**21**:216–29.
155. Lv Q, Chen G, Zhao L, et al. Mol2Context-vec: learning molecular representation from context awareness for drug discovery. *Brief Bioinform* 2021;**22**:bbab317.

156. Zhang Z, Cai J, Wang J. Duality-induced regularizer for tensor factorization based knowledge graph completion. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, Canada, p. 21604–15. NIPS, san diego, USA.
157. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;**11**: 333–7.
158. Guo S, Wang Y, Yuan H, et al. TAERT: triple-attentional explainable recommendation with temporal convolutional network. *Inform Sci* 2021;**567**:185–200.
159. Lin S, Wang Y, Zhang L, et al. MDF-SA-DDI: predicting drug-drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief Bioinform* 2022;**23**:bbab421.
160. Wu H, Pan X, Yang Y, Shen HB. Recognizing binding sites of poorly characterized RNA-binding proteins on circular RNAs using attention Siamese network. *Brief Bioinform* 2021;**22**:22.
161. Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics* 2019;**20**:415.
162. Chatr-Aryamontri A, Oughtred R, Boucher L, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 2017;**45**:D369–d379.
163. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–9.
164. Lin J, Wu L, Zhu J, et al. R2-DDI: relation-aware feature refinement for drug-drug interaction prediction. *Brief Bioinform* 2023;**24**:bbac576.
165. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. *Brief Bioinform* 2005;**6**:57–71.
166. Friedman C, Kra P, Yu H, et al. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics* 2001;**17**(Suppl 1):74–82.
167. Zhang Y, Wu HY, Xu J, et al. Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug-drug interactions from biomedical literature. *BMC Syst Biol* 2016;**10**(Suppl 3):67.
168. Wu HY, Karnik S, Subhadarshini A, et al. An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinformatics* 2013;**14**:35.
169. Tikk D, Palaga P, Leser U. A fast and effective dependency graph kernel for PPI relation extraction. *BMC Bioinformatics* 2010;**11**: 1–2.
170. Palmer M, Gildea D, Kingsbury P. The proposition Bank: an annotated corpus of semantic roles. *Comput Linguist* 2005;**31**: 71–106.
171. Airola A, Pyysalo S, Björne J, et al. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. *BMC Bioinformatics* 2008;**9** (Suppl 11):S2.
172. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 2010;**17**: 229–36.
173. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J Biomed Inform* 2003;**36**:462–77.
174. Zhang R, Cairelli MJ, Fiszman M, et al. Using semantic predications to uncover drug-drug interactions in clinical data. *J Biomed Inform* 2014;**49**:134–47.
175. Kilicoglu H, Shin D, Fiszman M, et al. SemMedDB: a PubMed-scale repository of biomedical semantic predications. *Bioinformatics* 2012;**28**:3158–60.
176. Seal RL, Braschi B, Gray K, et al. [Genenames.org](https://www.genenames.org): the HGNC resources in 2023. *Nucleic Acids Res* 2023;**51**:D1003–d1009.
177. Zheng W, Lin H, Luo L, et al. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinformatics* 2017;**18**:445.
178. Herrero-Zazo M, Segura-Bedmar I, Martínez P, Declerck T. The DDI corpus: an annotated corpus with pharmacological substances and drug-drug interactions. *J Biomed Inform* 2013;**46**: 914–20.
179. Zhou D, Miao L, He Y. Position-aware deep multi-task learning for drug-drug interaction extraction. *Artif Intell Med* 2018;**87**: 1–8.
180. Liu S, Tang B, Chen Q, Wang X. Drug-drug interaction extraction via convolutional neural networks. *Comput Math Methods Med* 2016;**2016**:1–8.
181. Huang D, Jiang Z, Zou L, Li L. Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. *Inform Sci* 2017;**415–416**: 100–9.
182. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;**9**:1735–80.
183. Bui QC, Sloot PM, van Mulligen EM, Kors JA. A novel feature-based approach to extract drug-drug interactions from biomedical text. *Bioinformatics* 2014;**30**:3365–71.
184. Sagae K, Tsujii JI. Dependency Parsing and Domain Adaptation with LR Models and Parser Ensembles. In: *EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Prague, Czech Republic. 2007, p. 1044–50. ACL, Stroudsburg, PA.
185. Jiang Z, Li L, Huang D et al. Training word embeddings for deep learning in biomedical text mining tasks. In: *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Washington DC, USA, 2015, p. 625–8. IEEE, New York, NY, USA.
186. Dou M, Ding J, Chen G, et al. IK-DDI: a novel framework based on instance position embedding and key external text for DDI extraction. *Brief Bioinform* 2023;**24**:bbad099.
187. Segura-Bedmar I. SemEval-2013 Task 9: extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013). In: *Joint Conference on Lexical and Computational Semantics*, Atlanta, USA, 2013, p. 341–50. ACL, Stroudsburg, PA.
188. Orkphol K, Yang W. Word sense disambiguation using cosine similarity collaborates with Word2vec and WordNet. *Future Internet* 2019;**11**:114.
189. He H, Chen G, Yu-Chian Chen C. 3DGT-DDI: 3D graph and text based neural network for drug-drug interaction prediction. *Brief Bioinform* 2022;**23**:bbac134.
190. Tosco P, Stiefl N, Landrum G. Bringing the MMFF force field to the RDKit: implementation and validation. *Journal of Cheminformatics* 2014;**6**:1–4.
191. Zhao Y, Wang CC, Chen X. Microbes and complex diseases: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbaa158.
192. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform* 2017;**70**:54–64.
193. Willett P. Similarity-based approaches to virtual screening. *Biochem Soc Trans* 2003;**31**:603–6.

194. Vilar S, Uriarte E, Santana L, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc* 2014;**9**:2147–63.
195. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 2002;**42**:1273–80.
196. Vilar S, Uriarte E, Santana L, et al. Detection of drug-drug interactions by modeling interaction profile fingerprints. *PloS One* 2013;**8**:e58321.
197. Liu M, Wu Y, Chen Y, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* 2012;**19**:e28–35.
198. Liu J, Gefen O, Ronin I, et al. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science* 2020;**367**:200–4.
199. Marroum PJ, Uppoor RS, Parmelee T, et al. In vivo drug-drug interaction studies- a survey of all new molecular entities approved from 1987 to 1997. *Clin Pharmacol Ther* 2000;**68**:280–5.
200. Zhang T, Leng J, Liu Y. Deep learning for drug-drug interaction extraction from the literature: a review. *Brief Bioinform* 2020;**21**:1609–27.
201. Lin X, Dai L, Zhou Y, et al. Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction. *Brief Bioinform* 2023;**24**:bbad235.