Briefings in Functional Genomics, 00(00), 2021, 1–11

https://doi.org/10.1093/bfgp/elab031
Review Paper

Current status and future prospects of drug–target interaction prediction

Xiaoqing Ru, Xiucai Ye, Tetsuya Sakurai, Quan Zou, Lei Xu and Chen Lin

Corresponding authors: Xiucai Ye, Department of Computer Science, University of Tsukuba, Tsukuba 3058577, Japan. Tel.: 029-853-5449; Fax: 029-853-5449; E-mail: yexiucai@cs.tsukuba.ac.jp, Chen Lin, School of Informatics, Xiamen University, Xiamen, Fujian, China , 361000. Tel.: 13606938620; E-mail: chenlin@xmu.edu.cn

Abstract

Drug–target interaction prediction is important for drug development and drug repurposing. Many computational methods have been proposed for drug–target interaction prediction due to their potential to the time and cost reduction. In this review, we introduce the molecular docking and machine learning-based methods, which have been widely applied to drug–target interaction prediction. Particularly, machine learning-based methods are divided into different types according to the data processing form and task type. For each type of method, we provide a specific description and propose some solutions to improve its capability. The knowledge of heterogeneous network and learning to rank are also summarized in this review. As far as we know, this is the first comprehensive review that summarizes the knowledge of heterogeneous network and learning to rank in the drug–target interaction prediction. Moreover, we propose three aspects that can be explored in depth for future research.

Key words: drug–target interaction; drug development; drug repurposing; machine learning

Introduction

Significance and current status of drug development

Drugs are compounds that can cause changes in the physiological functions of organs and cellular metabolic activities of living organisms after consumption, injection or absorption, and are important for maintaining human health [1, 2]. For various reasons, some drugs go through a process from flourishing to declining, i.e. a drug is no longer available for the prevention, treatment or diagnosis of a particular disease [3]. New drugs must be available to replace them. In addition, many diseases are still incurable or have no available drugs [4], and many new and complex diseases emerge every year. The development of drugs directly affects the quality and process of disease prevention and treatment [5], at present, it cannot catch up with the progress of diseases. Therefore, the development of effective drugs is urgent.

In general, the development of a new drug takes approximately 12–16 years and 1–2 billion dollars [6]. There are four phases of drug development: early drug discovery, preclinical trials, clinical trials and approval for marketing [7], with clinical trials divided into Phases I–III, which require consideration of how well the drug is tolerated by the organism, evaluation of the drug's efficacy and safety, and determination of therapeutic efficacy [8, 9]. Each of these phases takes a lot of time and money, which explains why drug development is characterized by high cost, high risk and long cycle. Currently, a lot of efforts

Xiaoqing Ru is a PhD student in the University of Tsukuba. Her research interest is drug–target interaction prediction and learning to rank.
Xiucai Ye is currently an assistant professor in the Department of Computer Science, and Center for Artificial Intelligence Research (C-AIR), University of Tsukuba. Her current research interests include feature selection, clustering, machine learning and bioinformatics.
Tetsuya Sakurai is currently a professor in the Department of Computer Science and is the director of the C-AIR, University of Tsukuba. His research interests include high-performance algorithms for large-scale simulations, data and image analysis, and deep neural network computations.
Quan Zou is a professor in University of Electronic Science and Technology of China. He is a senior member of IEEE and ACM. He majors in bioinformatics, machine learning and algorithms.
Lei Xu is an associate professor at the School of Electronic and Communication Engineering, Shenzhen Polytechnic. Her research interests include machine learning and bioinformatics.
Chen Lin is an associate professor in Xiamen University. She is a member of IEEE and ACM. Her research interests include data mining and recommendation system.

have been invested in drug development; the results are still not satisfactory [10]. According to Tufts CSDD impact Reports (https://csdd.tufts.edu/impact-reports), the US Food and Drug Administration approved only 126 cancer drugs for solid and blood tumors from 1980 to 2018. At Grakn Cosmos (https://www.grakncosmos.com/), the first global user conference organized by Grakn Labs in February 2020, Paul Agapow, the Director of Health Informatics at AstraZeneca, compared the drug development process to gambling—both have high investment and high risk. It is of great importance to reduce the high risk and investment of the 'gamble' of drug development by improving the efficiency at each stage.

Definition and significance of drug repurposing

Drug compounds may interact with other unexpected proteins in addition to the disease-related target proteins. Although this phenomenon is not expected by researchers, it has been proved that off-target drugs can produce unexpected new therapeutic effects [11]. For example, Imatinib Mesylate Capsules, which interacts with the Bcr-Abl fusion gene and is used to treat leukemia, has been shown to be effective in treating gastrointestinal stromal tumors [12]. Sildenafil, which is used to treat angina pectoris, has been shown to be effective in treating sexual dysfunction [13]. And Kinnings *et al.* [14] successfully applied the drugs tolcapone and entacapone for Parkinson's disease to anti-tuberculosis treatment by improving the predictive scoring of drug–target docking relationships. Finding new indications for existing drugs, i.e. drug repurposing [15], has attracted great attention from pharmaceutical companies, researchers, clinicians and even governments.

In recent decades, approximately 30% of new drug failures have been attributed to safety issues identified in clinical trials. COVID-19, which began at the end of 2019, has spread rapidly worldwide and its high infectiousness and insidiousness poses a serious threat to human health [16]. Obviously, it is not advisable to develop drugs for the treatment of this disease in accordance with the traditional process of new drug development [17]. Rapidly resolving the mechanism and predicting effective drugs are ways to deal with public health emergencies. At this point, systematic, large-scale screening of drugs that are already on the market and in clinical trials can be conducted. Most of the existing approved drugs are already guaranteed safe and have passed relevant pharmacological validation; repurposed drugs can enter the clinical phase more quickly and at lower cost than new drugs. Therefore, drug repurposing can significantly accelerate the drug development process and mitigate the impact of emergent diseases to a certain extent.

Of the 113 new drugs and biologics approved or released in 2017, only seven were completely new drugs (approved and released drugs with new mechanisms of action), whereas 36 were repurposed drugs [18]. Drug repurposing makes a drug available to patients within 3–12 years at an estimated total cost of 40–80 million dollars, a significant time and cost saving compared with new drug development [19, 20]. Drug repurposing promises to be an effective tool of treating diseases with high efficiency and low cost.

Necessity and significance of drug–target interaction prediction

The target of a drug is a biomolecule or biomolecular structure, usually a protein that binds specifically with a drug to produce a therapeutic effect on a disease [21, 22]. A protein that inhibits or promotes the occurrence and development of a disease can be considered as a candidate protein for the prevention and treatment of the disease. On the basis of this information, a relevant drug can be screened or developed for the prevention and treatment of the disease. Drug therapy is achieved when drug molecules bind to targets and regulate their biological activities, and the identification of drug–target interaction is beneficial to subsequent disease treatment. Therefore, investigating drug–target interactions is important for drug development and drug repurposing.

The purpose of predicting drug–target interactions is to identify the targets of new drugs and new targets of drugs [23–27]. With the development of molecular biology techniques and the completion of the Human Genome Project, a large number of proteins have been sequenced, but not all proteins are effective targets related to diseases. Moreover, known drug targets are only the tip of the iceberg compared with unknown drug targets [28]. In addition, many compounds have been synthesized over the past decades, such as the 109 million small molecules in the PubChem database, but the drug effects and targets of most of these compounds are unknown [29]. Among these compounds, there may be good drugs for treating diseases. These status quo make it urgent to explore drug–target interactions [30].

Two computational methods and their applications

Traditional experimental methods not only require a lot of human, material and financial resources, but also are easily affected by objective conditions. The use of biomedical experiments can also lead to many problems not being detected or dealt with in time because of efficiency issues; the rapid development of computer technology is an indispensable tool to solve these problems. Identifying drug–target interactions through computational methods can greatly reduce the broad search space for candidate drugs for downstream experimental verification, thereby significantly reducing the high cost and long cycle of developing new drugs [31–33]. Molecular docking, which has prevailed in the past few decades, is briefly introduced in this section. Recent years, molecular docking is arguably not considered generally as a mainstream method anymore for drug–target interaction predictions in the bioinformatics community, due to its well-known limitations. Machine learning methods can be used to process large-scale data, and with their rapid development, related machine learning methods have been applied in all stages of constructing drug–target interaction prediction models.

Molecular docking

Molecular docking is an important technology for computer-aided drug design [34–36]. Molecular docking places small molecules at the active site of the target protein and identifies the optimal conformation for the interaction of the small molecule (ligand) with the target macromolecule by continuously changing the ligand conformation, then predicts their binding mode and affinity. Dakshanamurthy *et al.* [37] performed molecular docking calculations using the crystal structures of human proteins and FDA-approved drugs, and identified an antiparasitic drug, mebendazole, with the potential to inhibit vascular endothelial growth factor receptor-2 (VEGFR-2).

At present, a number of easy-to-use and powerful docking methods and software have been developed [38–40], and molecular docking techniques allow a clear view of the binding of

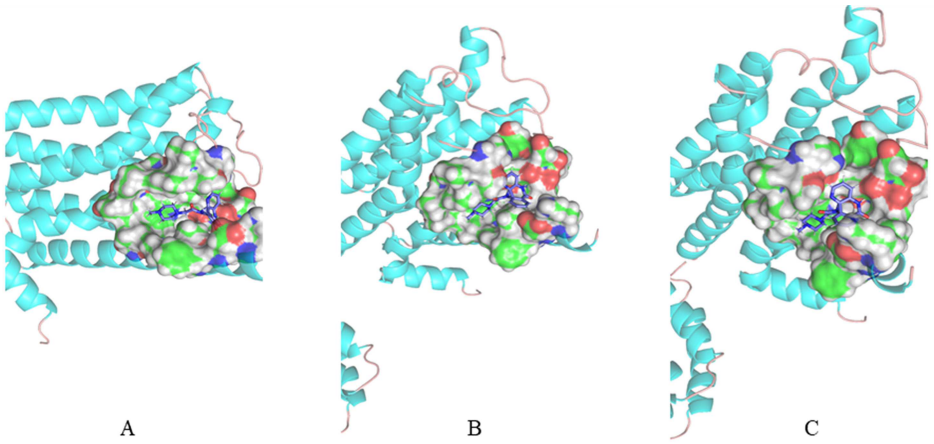

Figure 1. Example of molecular docking. Note that (**A**), (**B**), (**C**) are three presentation angles.

compounds to proteins. Figure 1 shows an example of docking (The docking process is completed by the swissdock (http://www.swissdock.ch/)) of pirenzepine (ligand) and muscarinic acetylcholine receptor M2, where the ligand is well embedded in the active pocket of the protein as observed from different angles. Because of these advantages, the molecular docking has become one of the common methods used to study interaction patterns between small and large molecules in the past [41, 42]. However, it also has some shortcomings, mostly because it is performed by computers to simulate the binding between ligand and receptor molecules. The ligand–receptor binding process is very complex and requires large and comprehensive sampling of all possible conformations to obtain an actual (or almost actual) binding conformation [7]. Conformational searching is required to find the optimal binding position and the simultaneous calculation of multiple conformations leads to a very large search space and computational effort [43], these processes can take a long time and have high computational cost, even when high-performance computers are used to screen individual targets. Current methods regard the docking as an independent process, and each docking is completely restarted, resulting in unnecessary waste of time and computational resources. In addition, molecular docking methods require the 3D structure of the protein is known, but there are many proteins whose 3D structures are unknown and not easily accessible, such as the GPCRs proteins [44, 45], which are important targets. Therefore, machine learning-based approaches to predict drug–target interactions are receiving much attention, and many studies have considered the problem from different perspectives to obtain high-performing drug–target interaction prediction models.

Machine learning

A common assumption in drug–target interaction prediction is that similar drugs target similar targets, and vice versa [46]. Sequences of the muscarinic acetylcholine receptors M1 (HSA:1128), M2 (HSA:1129), M3 (HSA:1131), M4 (HSA:1132) and M5 (HSA: 1133), are the most similar sequences to each other obtained by Blast comparison. And the parent components of pirenzepine (D08389), pirenzepine hydrochloride (D01297) and pirenzepine hydrochloride hydrate (D05276) are identical (note that the entry names of drugs and targets come from KEGG database (https://www.genome.jp/kegg/)). The relationship between them clearly validates the above hypothesis. The relationship between the five proteins and three compounds is shown in Figure 2. Most current applications are based on this assumption, and their performance is constantly improving. The binary interaction data sets for enzymes, ion channels, G-protein-coupled receptors (GPCRs) and nuclear receptors constructed by Yamanishi *et al.* [47] have been widely used in drug–target interaction studies. The existence of these 'gold standard' data sets provides a good reference for different methods and fully proves the progress of the methods. Many studies are based on these data sets, such as [48–53], and their methods have shown their respective advantages.

Existing applications can be divided into types from different perspectives. As shown in Figure 3, according to the processing form of drug and target data, these applications can be divided into feature-based, similarity-based, and network-based. According to the task type, applications can be divided into classification, regression and ranking.

I Types of methods according to the form of data processing

(i) Similarity-based method: The similarity-based approach is a direct manifestation of the hypothesis that 'similar drugs target similar targets and vice versa' [54, 55]. Figure 4A shows schematic diagram of the similarity-based method. It predicts the drug–target interaction through a score function. The higher the calculated score, the more likely the drug–target pair will interact. The similarity-based method can be divided into two perspectives of drug and target to calculate the score. For instance, in the drug perspective, Nearest neighbor (NN) [50] combines the similarity between the new drug and the known drug with the interaction between the known drug and the target to predict the interaction between the new drug and the target, and in the target perspective, NN uses the same principle to predict the interaction between the new target and the drug. And the final result is a combination of these two perspectives. In general, a large number of similarity-based methods have been proposed one after another, and these methods have their own characteristics. Bipartite local models (BLM) [50] are also based on the two perspectives of drugs and targets, but it turns the edge prediction problem into a well-known binary classification problem. Weighted nearest neighbor (WNN) [56] proposes a simple WNN algorithm, which uses the chemical and interaction information of known compounds to construct an interaction score profiles for a new drug. Keiser *et al.* [57] used a statistical-based chemical

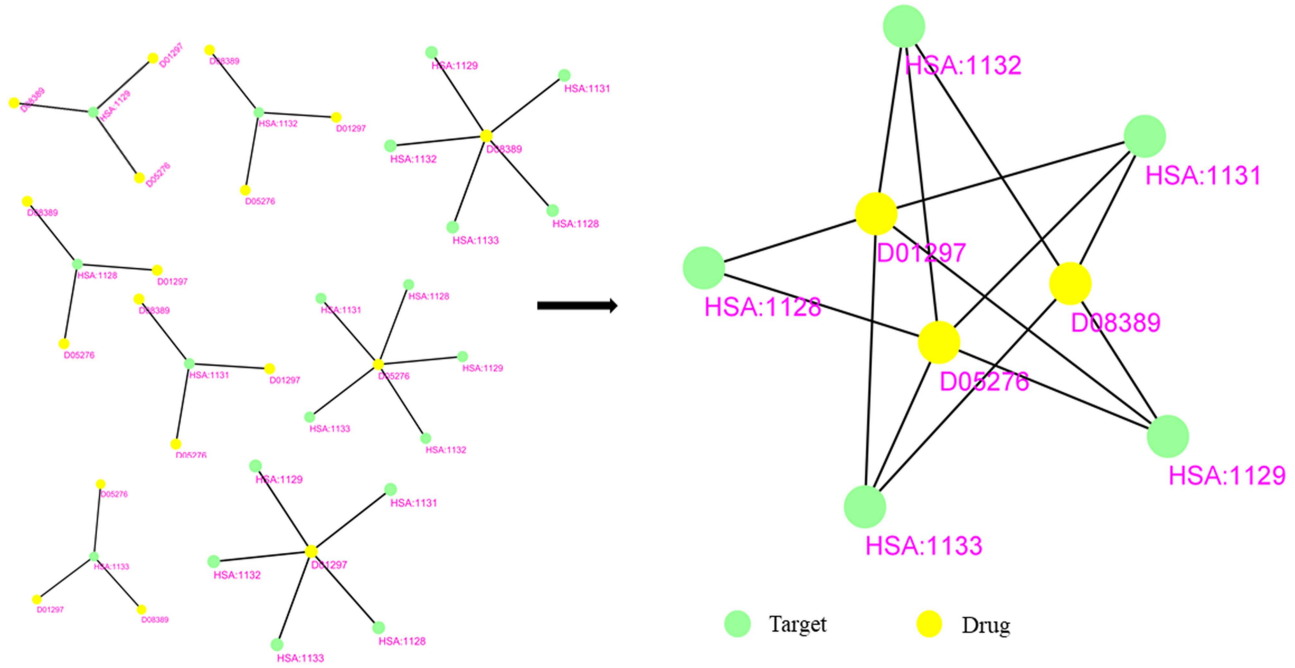

Figure 2. Explanation of the principle of similarity-based methods. Note that the entry names of drugs and targets come from KEGG database.

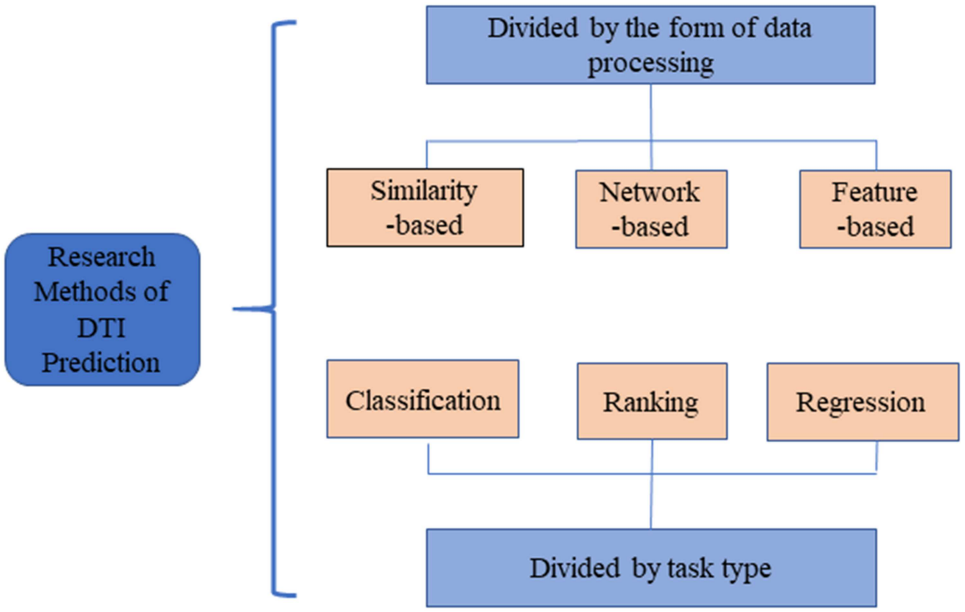

Figure 3. Types of specific applications. Note: DTI is an acronym for drug–target interaction.

method to predict new targets for small molecule drugs and drug compounds, and applied similarity coefficients to evaluate the 2D structural similarity of each drug and each target to identify new drug–target interactions. Zheng *et al.* [58] inferred drug–target relationships by synergistically processing drug similarity matrices, target similarity matrices and known interactions between the two. Hao *et al.* [59] improved the prediction performance by fusing drug and target similarity matrices and adding constraints on drug and target similarity matrices to calculate association relationships for drug–target pairs in the objective function.

(ii) Network-based method: In the network-based method, the nodes can represent drugs, targets, diseases, side effects, etc., and the edges can represent the relationship between them [60]. These networks are based on knowledge or use a variety of data resources for calculation and inference, and have various representations, such as drug–drug, target–target, drug–disease, disease–protein and so on. Figure 4B shows associations of heterogeneous networks. Network-based methods can be subdivided into two types: simple network-based and heterogeneous network-based.

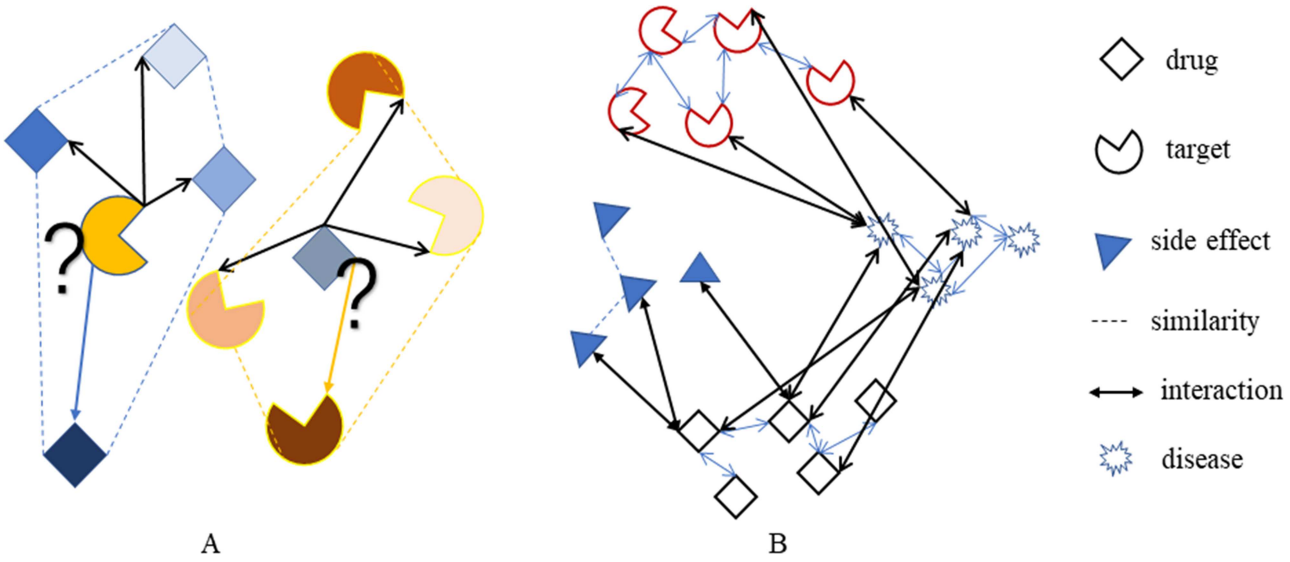

Figure 4. Schematic diagram of the similarity-based method (**A**) and associations of heterogeneous networks (**B**)

Simple network-based methods have relatively simple data sources. Most of them integrate drug structure, protein sequence and drug–target interaction network information, and only use one network topology. Such as Net Laplacian regularized least squares (NetLapRLS) [61], bipartite graph model [47] and Gaussian interaction profile (GIP). Many studies have shown that network-based methods are indeed effective methods for predicting drug–target interactions. GIP [49] defines the Gaussian interaction profile kernel based on the drug–target interaction profile, and through comparative experiments, it is verified that the performance of the GIP kernel is better than the kernel based on chemical and genomic information. Cheng *et al.* [48] applied drug-based similarity inference (DBSI), target-based similarity inference (TBSI) and network-based inference (NBI) to predict drug–target interaction, found that NBI performed best. In actuality, chemical structure, protein sequence information and other properties of them can be characterized by their various functional roles in biological systems, e.g. protein–protein interactions and drug–disease interactions, etc. For example, Campillos *et al.* [11] proposed to identify drug–target interactions based on the similarity of drug structure information and the similarity of drug side effects. Heterogeneous networks have high flexibility in modeling heterogeneous data, they have attracted widespread focus. NRWRH [62] integrated a chemical structure similarity network, a protein sequence similarity network, and the known drug–target interaction network into heterogeneous networks, and inferred potential drug–target relationship using a random walk with restart approach. DASPfind [63] also formed a heterogeneous network by combining a drug similarity network, a target protein similarity network, and a known drug–target protein bipartite graph network. It introduced an exponential decay function to fuse all pathways connecting a drug–target pair and finally predicted the drug–target interaction by traversing all simple pathways. DTINet [64] integrated diverse information from heterogeneous data sources (e.g. drugs, proteins, diseases and side-effects) by combining random walk with restart and diffusion component analysis, it found the best projection from drug space onto protein space, predicted the new drug–target interaction based on the geometric proximity of the mapping vector in the unified space and shown that incorporating additional network information can significantly improve the prediction accuracy. Some studies combined deep learning with heterogeneous networks, NeoDTI [60] integrated diverse information from heterogeneous network data and automatically learned topology-preserving representations of drugs and targets to facilitate drug–target interaction prediction. DeepDTnet [65] is a deep learning method based on a heterogeneous network embedded with 15 types of chemical, genomic, phenotypic and cellular network profiles to predict drug–target association relationships. Compared with other latest drug–target interaction prediction methods in the same period, NeoDTI and DeepDTnet both show superior identification capabilities. This phenomenon indicates that the combination of deep learning and heterogeneous networks for drug–target interaction prediction is an effective means to improve prediction performance.

(iii) Feature-based method: Feature-based method is widely used in the drug–target interaction prediction studies. This method needs to convert the drug and target data information into feature vectors. It is possible to extract drug and target features from various angles. E.g. Protein features can be extracted based on amino acid composition, pseudo-amino acid composition, amino acids physical and chemical properties, protein sequences evolution information and so on [66–70]. For drug–target interaction prediction, researchers use different feature extraction and processing methods to describe drug and target information. Tabei *et al.* [71] combined 881 compound structures of drugs and 876 Pfam domain structural information of target proteins using a tensor product approach. LRF-DTIs [72] used a pseudo position-specific scoring matrix to extract target information and applies FP2 molecular fingerprints to obtain drug information. iGPCR-Drug [73] performs discrete Fourier transform on drug molecular fingerprint to obtain drug features, and extracts GPCRs features based on pseudo-amino acid composition. Ru *et al.* [44] extracted protein features and drug features based on distance-based Top-n-gram algorithm and general descriptors of compounds. Each compound can be represented by a SMILE sequence, the drug features can also be extracted based on this sequence. For example, Hirohara *et al.* [74] converted

the SMILES string into a two-dimensional matrix and used a convolutional neural network (CNN) to extract its features.

II Types of methods according to the form of task type

Existing drug–target interaction prediction studies can be divided into two types depending on the prediction form [75–79]: one is to explore whether the drug can interact with the target, and the other is to explore the degree of drug–target interaction. Exploring whether the drug can interact with the target is a binary classification task, and exploring the degree of drug–target interaction can be regarded as a regression or ranking task. Figure 5 shows processes of these tasks.

(i) Classification task: Most current studies take drug–target interaction prediction as a supervised binary classification problem. Researchers have introduced a variety of classification algorithms to obtain better performance. E.g. Pred-binding [80] extracts the molecular structure and protein sequence features, and uses support vector machines and random forests to classify the docking of drugs and targets. iDTI-ESBoost [81] uses the AdaBoost algorithm to classify after extracting features based on the drug molecular fingerprint information and the target evolution and structure information. Binary classification lays a good foundation for the initial stage of drug development, but in order to further accelerate the process of drug development, it is far from enough to explore whether the drug can interact with the target.

(ii) Ranking task: Exploring the strength of drug–target interactions, and finding a target (or drug) that has a strong interaction with the drug (or target) can effectively reduce the number of downstream biomedical verification experiments, thereby achieving the effect of accelerating process and saving cost. Drug–target interaction prediction has been considered as a ranking task, and learning to rank, which is widely used in information retrieval, is applied to solve such task. Zhang *et al.* [82] represented the drug by general descriptors, and represented the target information by the amino acid composition, transformation, and distribution, and then used learning to rank to simultaneously learn the information under different experimental conditions and across targets. In addition to extracting features according to protein sequence information and compound descriptors, DrugE-Rank [83] also extracted features according to the prediction results of six classifiers, finally produced drug ranking results by inputting these features into the learning to rank algorithm. Ru *et al.* [44] used learning to rank algorithm to explore the interaction between drugs and GPCRs.

(iii) Regression task: Learning to rank algorithm outputs relative relevance of query and document. Therefore, it only can obtain approximate ranking of the target (or drug) that interacts with the drug (or target). To get a continuum of binding strength values, deep learning approaches have been applied to protein–ligand interaction scoring. E.g. DeepDTA [84] uses CNN to model protein sequences and compound 1D representations, and uses fully connected layers in affinity prediction task. GraphDTA [85] represents drugs as graphs and uses graph neural networks to predict drug–target affinity. The procedure of using deep learning to solve the drug–target interaction prediction problem can be summarized as shown in Figure 5B.

Discussion

Compared with molecular docking, one obvious advantage of machine learning-based approaches is that it does not require protein 3D structural information. At present, machine learning-based methods have been widely used in drug–target interaction prediction. This review divides these methods according to the data processing form and task type. Therefore, this chapter analyzes and summarizes the characteristics of these methods from different types, and proposes solutions to some existing problems.

(i) The principle of similarity-based methods is simple and easy to understand. In process, it does not involve complex procedures of feature extraction and feature selection. In the future, similarity-based methods may be used as the underlying technology to obtain useful information and be combined with other types of methods to better address drug–target interaction prediction problems. E.g. DrugE-Rank [83] used six cutting-edge similarity-based methods as component methods, and applied learning to rank algorithm to predict drug–target interactions. It also shown that each similarity method has its own uniqueness and focuses on different points. Therefore, integration-based models that combine multiple similarities may provide more accurate results. In principle, similarity-based method is a direct manifestation of the hypothesis that 'similar drugs target similar targets and vice versa', but in some special cases, this hypothesis is not convincing. Yamanishi *et al.* [47] pointed out that some low similarity target proteins in enzyme can bind to similar drugs.

(ii) Network is a simple data structure. The nodes in the network have rich attribute information. There are complex information networks among them, and different associations can be found through statistics and calculation methods [86, 87]. Compared with a homogeneous network that only considers one kind of node and one topological relationship, a heterogeneous network can fuse more information and effectively embed the rich structural and semantic information into a low-dimensional representation [46, 88]. At present, network-based methods have not been widely used in drug–target interaction prediction, because both homogeneous and heterogeneous networks require a lot of prior knowledge, which is difficult to obtain.

(iii) Feature engineering plays a decisive role in machine learning [89]. Extracting features from one angle cannot accurately and comprehensively describe drug or target information, and simple splicing cannot mine deeper information. These two are common phenomena for extracting and processing features in drug–target interaction prediction research. To address the feature engineering problem, studies should focus on different perspectives to extract drug and target features. There may be some redundancy among the features extracted by these different perspectives; therefore, effective feature processing is required. For example, Ru *et al.* [44] verified the positive impact of principal component analysis (PCA) on their research. The dimensionality of the feature set was reduced after PCA processing, but the performance of the model built based on this feature set was improved. This step is currently overlooked in many drug–target interaction prediction studies. In addition, it is also necessary to consider the relationship between drug and target features, deep learning and multi-view learning may be effective tools to explore these relationships [90].

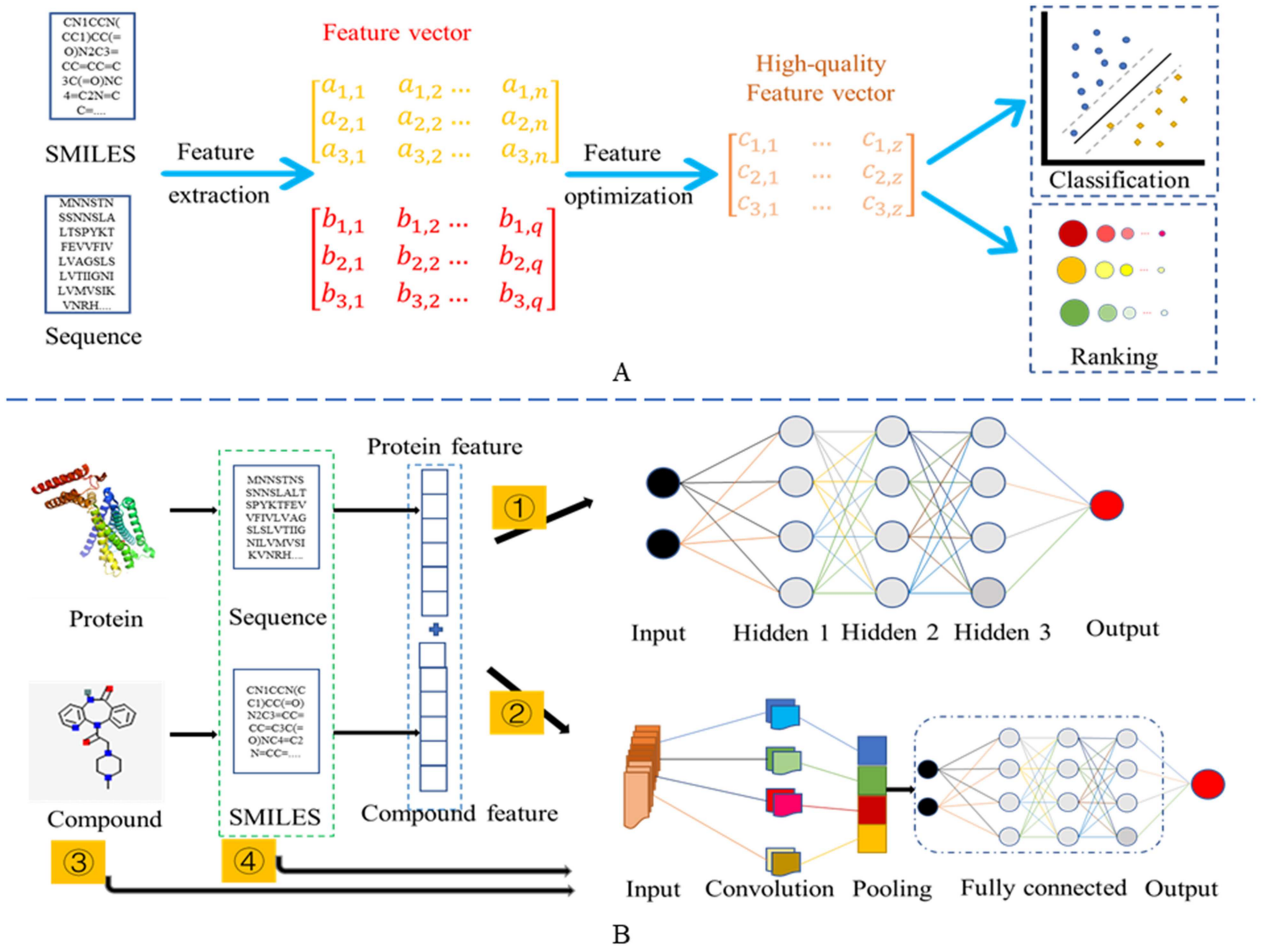

Figure 5. Processes of classification task, ranking task (**A**) and regression task (**B**). In (**B**), ①–④ represent four forms of information input into the deep learning network. Specifically, forms ① and ② both extract features based on protein sequence and drug SMILES, and then input these features into the neural network. The differences between them are: ① directly inputs information into the fully connected neural network (DNN) for prediction, and ② uses the CNN to process the features and then inputs them into DNN for prediction. Form ③ is to directly use the structural information of proteins and drugs as the input of the deep learning framework. Form ④ is to directly use the protein sequence and drug SMILES information as the input. Note that the structure diagram of the compound in Part B comes from PubChem.

(iv) Building classification models require four procedures: dataset acquisition, feature extraction, feature optimization and optimal classification algorithm selection [91–96]. High-performing models can be obtained by continuously optimizing these steps. These indicate that more progress can still be made in classification models. It is important to consider the credibility of the data when building classification models. Predicting drug–target interactions using supervised classification methods requires a certain number of positive and negative samples, which have high reliability through experimental validation. At present, only a small number of experimentally verified samples are disclosed, and some negative examples used in many studies have not been experimentally verified, and it cannot be determined whether they are true negative examples or unverified potential positive examples. Moreover, the number of unverified drug–target pairs is much larger than the number of verified drug–target pairs, which leads to data imbalance problem. Data credibility can only be verified by the accumulation of time and experience, but data imbalance problem can be alleviated with existing technologies. E.g. Both Wang *et al.* [97] and Wang *et al.* [98] randomly selected negative samples to ensure data balance. Pdti-EssB [99] uses random under-sampling and under-sampling clustering. iDTI-ESBoost [81] uses a novel data balancing technology—cluster based under sampling. NetLapRLS [61] and NormMulInf [100] directly treat drug–target interaction prediction as a semi-supervised problem. In general, data imbalance problem is a key issue that requires continuous attention.

(v) Learning to rank was originally applied in information retrieval. Its principle is similar to the process of searching information on the World Wide Web [101], that is, the user enters a query, the search engine outputs related documents, and these documents are ranked in descending order of relevance. Using the uniqueness of learning to rank can solve multiple problems in drug–target interaction prediction: Considering drug–target interaction prediction as a ranking task successfully circumvents the requirement of negative samples in classification methods and also enables to explore which targets have strong interaction with a drug. As the number of complex diseases increases year by year, it is imperative to develop effective multi-target drugs. There is a one-to-many relationship between

query and document in learning to rank, which can be used to predict multi-target drugs. Therefore, it is valuable to apply learning to rank in drug–target interaction prediction. It is also necessary to pay attention to data imbalance problem when constructing ranking models, and try to ensure that the number of documents corresponding to each query is balanced. Finally, it should be noted that the output of learning to rank is relative relevance, that is, this method focuses on order, rather than specific and accurate relevance.

(vi) Regression tasks can derive specific values, and most of these tasks rely on deep learning frameworks. The advantage of deep learning lies in the ability to learn complex relationships between input features and output decisions from large-scale data [102–110]. Deep learning for exploring drug–target interactions is still in its infancy, it has already shown great potential. But this does not mean that this method is perfect. Deep learning frameworks usually have many parameters and require multiple optimizations to train these parameters. Moreover, these frameworks need a long execution time, and may also cause over-fitting results.

Conclusion and outlook

In this review, we focus on the applications of machine learning in drug–target interaction prediction. According to the processing form of drug and target data, the existing studies are divided into three types: similarity-based, network-based and feature-based. And according to the task type, they are divided into three types: classification, regression and ranking. Then, we summarize and analyze the characteristics, shortcomings and aspects that can be further improved of each type of method. This review will provide a good guide for future work on drug–target interaction prediction.

The existing studies not only have achieved good performances, but also have focused on in-depth and specific problems, which are beneficial for drug development and drug repurposing. For future work, we propose the following ideas.

(i) Few of the existing drug–target interaction prediction studies have considered drug–drug interactions. The number of drug combinations per capita in conventional medical care is approximately five [111]. Summarizing the experience in clinical pharmacology, it concluded that drug combinations may enhance or counteract the efficacy, or may produce side effect that is detrimental to disease treatment [112]. Moreover, adverse reactions caused by drug–drug interactions are an important reason for the withdrawal of drugs from the market [113]. Therefore, more attention should be paid to drug–drug interactions in future drug–target interaction prediction studies.

(ii) Although the vast majority of targets are proteins, the interactions between drugs and small molecules [114, 115] (e.g. miRNAs [116]) also should be considered in the drug development process for better treatment of complex diseases.

(iii) Each method has its own advantages and disadvantages. It is possible to improve prediction performance by combining different methods. The drug and target information obtained through similarity-based methods and network-based methods can be input as features into classification, ranking or regression algorithms, which may improve the performance of the models. For example, NeoDTI and DeepDTnet models constructed by combining deep learning methods and network methods, and DrugE-Rank model combining similarity methods and ranking methods, all show excellent performance. Similarly, other different combinations of these methods may also lead to good performance.

Key Points

- Computer technologies have been widely used in the drug–target interaction prediction. Particularly, machine learning has attracted intensive interest.
- The specific applications, which are predicting drug–target interaction through machine learning, can be divided into multiple types according to the data processing form and task type.
- Different types of methods have their own characteristics. These methods focus on different content and have different steps to deal with problems.
- For future work, this review concludes that combining different methods is an effective means to improve the prediction performance of drug–target interactions.

Funding

New Energy and Industrial Technology Development Organization 265 (NEDO) and JST COI-NEXT; Grant of Shenzhen Polytechnic (No. 6021310015 K).

Conflict of interest

All authors declare no conflict of interest.

References

1. De Clercq E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* 2009;**33**(4):307–20.
2. Muzzio FJ, Shinbrot T, Glasser BJ. *Powder Technology in the Pharmaceutical Industry: the Need to Catch Up Fast*. Elsevier, 2002.
3. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet* 2000;**356**(9237):1255–9.
4. Schneider LS, Mangialasche F, Andreasen N, *et al*. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. *J Intern Med* 2014;**275**(3):251–83.
5. Liao Y-D, Jiang Z-R. MoABank: an integrated database for drug mode of action knowledge. *Curr Bioinform* 2019;**14**(5):446–9.
6. Lotfi Shahreza M, Ghadiri N, Mousavi SR, *et al*. A review of network-based approaches to drug repositioning. *Brief Bioinform* 2018;**19**(5):878–92.
7. D'Souza S, Prema K, Balaji S. Machine learning models for drug–target interactions: current knowledge and future directions. *Drug Discov Today* 2020;**25**(4):748–56.
8. Mahan VL. Clinical trial phases. *Int J Clin Med* 2014;**5**(21):1374–83.
9. Stephenson N, Shane E, Chase J, *et al*. Survey of machine learning techniques in drug discovery. *Curr Drug Metab* 2019;**20**(3):185–93.

10. Esch EW, Bahinski A, Huh D. Organs-on-chips at the frontiers of drug discovery. *Nat Rev Drug Discov* 2015;**14**(4):248–60.
11. Campillos M, Kuhn M, Gavin A-C, *et al*. Drug target identification using side-effect similarity. *Science* 2008;**321**(5886):263–6.
12. Dagher R, Cohen M, Williams G, *et al*. Approval summary: imatinib mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors. *Clin Cancer Res* 2002;**8**(10):3034–8.
13. Kaplan SA, Reis RB, Kohn IJ, *et al*. Safety and efficacy of sildenafil in postmenopausal women with sexual dysfunction. *Urology* 1999;**53**(3):481–6.
14. Kinnings SL, Liu N, Tonge PJ, *et al*. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J Chem Inf Model* 2011;**51**(2):408–19.
15. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, *et al*. Drug repositioning and repurposing: terminology and definitions in literature. *Drug Discov Today* 2015;**20**(8):1027–34.
16. Cheng L, Han X, Zhu Z, *et al*. Functional alterations caused by mutations reflect evolutionary trends of SARS-CoV-2. *Brief Bioinform* 2021;**22**(2):1442–50.
17. Chong CR, Sullivan DJ. New uses for old drugs. *Nature* 2007;**448**(7154):645–6.
18. Graul AI, Cruces E, Stringer M. The year's new drugs & biologics, 2013: part I. *Drugs Today (Barc)* 2014;**50**(1):51–100.
19. Hurle M, Yang L, Xie Q, *et al*. Computational drug repositioning: from data to therapeutics. *Clin Pharmacol Ther* 2013;**93**(4):335–41.
20. Papapetropoulos A, Szabo C. Inventing new therapies without reinventing the wheel: the power of drug repurposing. *Br J Pharmacol* 2018;**175**(2):165.
21. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov* 2002;**1**(9):727–30.
22. Munir A, Malik SI, Malik KA. Proteome mining for the identification of putative drug targets for human pathogen *Clostridium tetani*. *Curr Bioinform* 2019;**14**(6):532–40.
23. Ding Y, Tang J, Guo F. Identification of drug–target interactions via fuzzy bipartite local model. *Neural Comput Applic* 2020;**32**(14):10303–19.
24. Ding Y, Tang J, Guo F. Identification of drug–target interactions via dual Laplacian regularized least squares with multiple kernel fusion. *Knowl-Based Syst* 2020;**204**:106254.
25. Yu L, Wang M, Yang Y, *et al*. Predicting therapeutic drugs for hepatocellular carcinoma based on tissue-specific pathways. *PLoS Comput Biol* 2021;**17**(2):e1008696.
26. Wang J, Wang H, Wang X, *et al*. Predicting drug-target interactions via FM-DNN learning. *Current Bioinformatics* 2020;**15**(1):68–76.
27. Wang J, Shi Y, Wang X, *et al*. A drug target interaction prediction based on LINE-RF learning. *Curr Bioinform* 2020;**15**(7):750–7.
28. Landry Y, Gies JP. Drugs and their molecular targets: an updated overview. *Fundam Clin Pharmacol* 2008;**22**(1):1–18.
29. Sayers EW, Barrett T, Benson DA, *et al*. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2010;**39**(suppl_1):D38–51.
30. Liu J, Lian X, Liu F, *et al*. Identification of novel key targets and candidate drugs in oral squamous cell carcinoma. *Curr Bioinform* 2020;**15**(4):328–37.
31. Kitchen DB, Decornez H, Furr JR, *et al*. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;**3**(11):935–49.
32. Peska L, Buza K, Koller J. Drug-target interaction prediction: a Bayesian ranking approach. *Comput Methods Programs Biomed* 2017;**152**:15–21.
33. Ding Y, Tang J, Guo F. Identification of drug-target interactions via multiple information integration. *Inform Sci* 2017;**418**:546–60.
34. Jakhar R, Dangi M, Khichi A, *et al*. Relevance of molecular docking studies in drug designing. *Curr Bioinform* 2020;**15**(4):270–8.
35. Zulfiqar H, Masoud MS, Yang H, *et al*. Screening of prospective plant compounds as H1R and CL1R inhibitors and its antiallergic efficacy through molecular docking approach. *Comput Math Methods Med* 2021;**2021**:1–9.
36. Srivastava N, Mishra BN, Srivastava P. In-silico identification of drug lead molecule against pesticide exposed-neurodevelopmental disorders through network-based computational model approach. *Curr Bioinform* 2019;**14**(5):460–7.
37. Dakshanamurthy S, Issa NT, Assefnia S, *et al*. Predicting new indications for approved drugs using a proteochemometric method. *J Med Chem* 2012;**55**(15):6832–48.
38. Damm-Ganamet KL, Smith RD, Dunbar JB, *et al*. CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J Chem Inf Model* 2013;**53**(8):1853–70.
39. van Zundert GCP, Rodrigues J, Trellet M, *et al*. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* 2016;**428**(4):720–5.
40. Morris GM, Huey R, Lindstrom W, *et al*. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;**30**(16):2785–91.
41. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev* 2017;**9**(2):91–102.
42. Roy S, Kumar A, Baig MH, *et al*. Virtual screening, ADMET profiling, molecular docking and dynamics approaches to search for potent selective natural molecules based inhibitors against metallothionein-III to study Alzheimer's disease. *Methods* 2015;**83**:105–10.
43. Ding H, Takigawa I, Mamitsuka H, *et al*. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Brief Bioinform* 2014;**15**(5):734–47.
44. Ru X, Wang L, Li L, *et al*. Exploration of the correlation between GPCRs and drugs based on a learning to rank algorithm. *Comput Biol Med* 2020;**119**:103660.
45. Cooke RM, Brown AJ, Marshall FH, *et al*. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug Discov Today* 2015;**20**(11):1355–64.
46. Luo Y, Zhao X, Zhou J, *et al*. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):1–13.
47. Yamanishi Y, Araki M, Gutteridge A, *et al*. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**(13):i232–40.
48. Cheng F, Liu C, Jiang J, *et al*. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012;**8**(5):e1002503.
49. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* 2011;**27**(21):3036–43.

50. Bleakley K, Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* 2009;**25**(18):2397–403.
51. Yamanishi Y, Kotera M, Kanehisa M, *et al*. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 2010;**26**(12):i246–54.
52. Chen X, Liu M-X, Yan G-Y. Drug–target interaction prediction by random walk on the heterogeneous network. *Mol Biosyst* 2012;**8**(7):1970–8.
53. Pahikkala T, Airola A, Pietilä S, *et al*. Toward more realistic drug–target interaction predictions. *Brief Bioinform* 2015;**16**(2):325–37.
54. Klabunde T. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br J Pharmacol* 2007;**152**(1):5–7.
55. Schuffenhauer A, Floersheim P, Acklin P, *et al*. Similarity metrics for ligands reflecting the similarity of the target proteins. *J Chem Inf Comput Sci* 2003;**43**(2):391–405.
56. Van Laarhoven T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS One* 2013;**8**(6):e66952.
57. Keiser MJ, Setola V, Irwin JJ, *et al*. Predicting new molecular targets for known drugs. *Nature* 2009;**462**(7270):175–81.
58. Zheng X, Ding H, Mamitsuka H, *et al*. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.
59. Hao M, Bryant SH, Wang Y. Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Sci Rep* 2017;**7**(1):1–11.
60. Wan F, Hong L, Xiao A, *et al*. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* 2019;**35**(1):104–11.
61. Xia Z, Wu L-Y, Zhou X, *et al*. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst Biol* 2010 BioMed Central.;**4**(Suppl 2):S6.
62. Chen X, Yan G-Y. Nrwrh for drug target prediction. *Comput Syst Biol* 2010;**13**:219–26.
63. Ba-Alawi W, Soufan O, Essack M, *et al*. DASPfind: new efficient method to predict drug–target interactions. *J Chem* 2016;**8**(1):1–9.
64. Luo Y, Zhao X, Zhou J, *et al*. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):573.
65. Zeng X, Zhu S, Lu W, *et al*. Target identification among known drugs by deep learning from heterogeneous networks. *Chem Sci* 2020;**11**(7):1775–97.
66. Liu B, Gao X, Zhang H. BioSeq-Analysis2. 0: an updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. *Nucleic Acids Res* 2019;**47**(20): e127–7.
67. Tang Y-J, Pang Y-H, Liu B. IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning. *Bioinformatics* 2020;**36**(21): 5177–5186.
68. Liu B, Zhu Y, Yan K. Fold-LTR-TCP: protein fold recognition based on triadic closure principle. *Brief Bioinform* 2020;**21**(6):2185–93.
69. Wei L, Tang J, Zou Q. Local-DPP: an improved DNA-binding protein prediction method by exploring local evolutionary information. *Inform Sci* 2017;**384**:135–44.
70. Wang H, Ding Y, Tang J, *et al*. Identification of membrane protein types via multivariate information fusion with Hilbert–Schmidt independence criterion. *Neurocomputing* 2020;**383**:257–69.
71. Tabei Y, Pauwels E, Stoven V, *et al*. Identification of chemogenomic features from drug–target interaction networks using interpretable classifiers. *Bioinformatics* 2012;**28**(18):i487–94.
72. Shi H, Liu S, Chen J, *et al*. Predicting drug-target interactions using lasso with random forest based on evolutionary information and chemical structure. *Genomics* 2019;**111**(6):1839–52.
73. Xiao X, Min J-L, Wang P, *et al*. iGPCR-drug: a web server for predicting interaction between GPCRs and drugs in cellular networking. *PLoS One* 2013;**8**(8):e72234.
74. Hirohara M, Saito Y, Koda Y, *et al*. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. *BMC Bioinformatics* 2018;**19**(19):83–94.
75. Zhang Z-Y, Yang Y-H, Ding H, *et al*. Design powerful predictor for mRNA subcellular location prediction in Homo sapiens. *Brief Bioinform* 2021;**22**(1):526–35.
76. Yang H, Luo Y, Ren X, *et al*. Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators. *Inform Fusion* 2021;**75**:140–9.
77. Yu L, Shi Y, Zou Q, *et al*. Exploring drug treatment patterns based on the action of drug and multilayer network model. *Int J Mol Sci* 2020;**21**(14):5014.
78. Yu L, Xu F, Gao L. Predict new therapeutic drugs for hepatocellular carcinoma based on gene mutation and expression. *Front Bioeng Biotechnol* 2020;**8**:8.
79. Zhao X, Jiao Q, Li H, *et al*. ECFS-DEA: an ensemble classifier-based feature selection for differential expression analysis on expression profiles. *BMC Bioinformatics* 2020;**21**(1): 43.
80. Shar PA, Tao W, Gao S, *et al*. Pred-binding: large-scale protein–ligand binding affinity prediction. *J Enzyme Inhib Med Chem* 2016;**31**(6):1443–50.
81. Rayhan F, Ahmed S, Shatabda S, *et al*. iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting. *Sci Rep* 2017;**7**(1):1–18.
82. Zhang W, Ji L, Chen Y, *et al*. When drug discovery meets web search: learning to rank for ligand-based virtual screening. *J Chem* 2015;**7**(1):1–13.
83. Yuan Q, Gao J, Wu D, *et al*. DrugE-rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics* 2016;**32**(12):i18–27.
84. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;**34**(17):i821–9.
85. Nguyen T, Le H, Quinn T P, *et al*. GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics (Oxford, England)* 2021;**37**(8):1140–1147.
86. Jin S, Zeng X, Fang J, *et al*. A network-based approach to uncover microRNA-mediated disease comorbidities and potential pathobiological implications. *NPJ Syst Biol Appl* 2019;**5**(1):1–11.
87. Ding Y, Tang J, Guo F. Identification of drug-side effect association via multiple information integration with centered kernel alignment. *Neurocomputing* 2019;**325**:211–24.

88. Zhao X, Chen L, Guo Z-H, *et al*. Predicting drug side effects with compact integration of heterogeneous networks. *Curr Bioinform* 2019;**14**(8):709–20.
89. Wei L, Xing P, Shi G, *et al*. Fast prediction of protein methylation sites using a sequence-based feature selection technique. *IEEE/ACM Trans Comput Biol Bioinform* 2017;**16**(4):1264–73.
90. Jin Q, Meng Z, Pham TD, *et al*. DUNet: a deformable network for retinal vessel segmentation. *Knowl-Based Syst* 2019;**178**:149–62.
91. Shao J, Yan K, Liu B. FoldRec-C2C: protein fold recognition by combining cluster-to-cluster model and protein similarity network. *Brief Bioinform* 2021;**22**(3): bbaa144.
92. Cheng L, Zhao H, Wang P, *et al*. Computational methods for identifying similar diseases. *Mol Ther-Nucl Acids* 2019;**18**:590–604.
93. Su R, Hu J, Zou Q, *et al*. Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools. *Brief Bioinform* 2020;**21**(2):408–20.
94. Wei L, Chen H, Su R. M6APred-EL: a sequence-based predictor for identifying N6-methyladenosine sites using ensemble learning. *Mol Ther-Nucl Acids* 2018;**12**:635–44.
95. Wei L, Wan S, Guo J, *et al*. A novel hierarchical selective ensemble classifier with bioinformatics application. *Artif Intell Med* 2017;**83**:82–90.
96. Dong L, Wang J, Wang G. BYASE: a python library for estimating gene and isoform level allele-specific expression. *Bioinformatics* 2020;**36**(19):4955–6.
97. Wang Y-C, Yang Z-X, Wang Y, *et al*. Computationally probing drug-protein interactions via support vector machine. *Lett Drug Design Discov* 2010;**7**(5):370–8.
98. Wang L, You Z-H, Chen X, *et al*. A computational-based method for predicting drug–target interactions by using stacked autoencoder deep neural network. *J Comput Biol* 2018;**25**(3):361–73.
99. Mahmud SH, Chen W, Meng H, *et al*. Prediction of drug-target interaction based on protein features using undersampling and feature selection techniques with boosting. *Anal Biochem* 2020;**589**:113507.
100. Peng L, Liao B, Zhu W, *et al*. Predicting drug–target interactions with multi-information fusion. *IEEE J Biomed Health Inform* 2015;**21**(2):561–72.
101. Ru X, Ye X, Sakurai T, *et al*. Application of learning to rank in bioinformatics tasks. *Brief Bioinform* 2021. doi: 10.1093/bib/bbaa394.
102. Zhao T, Hu Y, Peng J, *et al*. DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 2020;**36**(16):4466–72.
103. Cheng L. Computational and biological methods for gene therapy. *Curr Gene Ther* 2019;**19**(4):210.
104. Wang D, Zhang Z, Jiang Y, *et al*. DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res* 2021;**49**(8):e46.
105. Lv H, Dao F-Y, Guan Z-X, *et al*. Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. *Brief Bioinform* 2020. doi: 10.1093/bib/bbaa255.
106. Yu L, Zhou D, Gao L, *et al*. Prediction of drug response in multilayer networks based on fusion of multiomics data. *Methods* 2020. doi: 10.1016/j.ymeth.2020.08.006.
107. Xu H, Zeng W, Zeng X, *et al*. A polar-metric-based evolutionary algorithm. *IEEE Trans Cyber* 2020;1–12. doi: 10.1109/TCYB.2020.2965230.
108. Zeng X, Song X, Ma T, *et al*. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *J Proteome Res* 2020;**19**(11):4624–36.
109. Liu Y, Huang Y, Wang G, *et al*. A deep learning approach for filtering structural variants in short read sequencing data. *Brief Bioinform* 2020. doi: 10.1093/bib/bbaa370.
110. Wei L, Ding Y, Su R, *et al*. Prediction of human protein subcellular localization using deep learning. *J Paral Distrib Comput* 2018;**117**:212–7.
111. Huang JL, Niu CQ, Green CD, *et al*. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Comput Biol* 2013;**9**(3):e1002998.
112. Lin X, Quan Z, Wang Z-J, *et al*. Kgnn: knowledge graph neural network for drug-drug interaction prediction. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20 (International Joint Conferences on Artificial Intelligence Organization)*. 2020.
113. Dhingra NK, Colombino M, Jovanovic MR. Structured decentralized control of positive systems with applications to combination drug therapy and leader selection in directed networks. *IEEE Trans Control Netw Syst* 2019;**6**(1): 352–62.
114. Wang H, Tang J, Ding Y, *et al*. Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment. *Brief Bioinform* 2021. doi: 10.1093/bib/bbaa409.
115. Zeng X, Zhong Y, Lin W, *et al*. Predicting disease-associated circular RNAs using deep forests combined with positive-unlabeled learning methods. *Brief Bioinform* 2020;**21**(4):1425–36.
116. Zeng X, Liu L, Lü L, *et al*. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 2018;**34**(14):2425–32.