Briefings in Bioinformatics, 22(1), 2021, 451–462

doi: 10.1093/bib/bbz152
Advance Access Publication Date: 23 December 2019
Problem Solving Protocol

DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features

Yanyi Chu, Aman Chandra Kaushik, Xiangeng Wang, Wei Wang, Yufang Zhang, Xiaoqi Shan, Dennis Russell Salahub, Yi Xiong and Dong-Qing Wei

Corresponding authors. Yi Xiong and Dong-Qing Wei, State Key Laboratory of Microbial Metabolism, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nanshan District, Shenzhen, Guangdong, 518055, China. Tel.: +86 21-34204573; Email: xiongyi@sjtu.edu.cn, dqwei@sjtu.edu.cn

Abstract

Drug–target interactions (DTIs) play a crucial role in target-based drug discovery and development. Computational prediction of DTIs can effectively complement experimental wet-lab techniques for the identification of DTIs, which are typically time- and resource-consuming. However, the performances of the current DTI prediction approaches suffer from a problem of low precision and high false-positive rate. In this study, we aim to develop a novel DTI prediction method for improving the prediction performance based on a cascade deep forest (CDF) model, named DTI-CDF, with multiple similarity-based features between drugs and the similarity-based features between target proteins extracted from the heterogeneous graph, which contains known DTIs. In the experiments, we built five replicates of 10-fold cross-validation under three different experimental settings of data sets, namely, corresponding DTI values of certain drugs (S_D), targets (S_T), or drug-target pairs (S_P) in the training sets are missed but existed in the test sets. The experimental results demonstrate that our proposed approach DTI-CDF achieves a significantly higher performance than that of the traditional ensemble learning-based methods such as random forest and XGBoost, deep neural network, and the state-of-the-art methods such as DDR. Furthermore, there are 1352 newly predicted DTIs which are proved to be correct by KEGG and DrugBank databases. The data sets and source code are freely available at https://github.com//a96123155/DTI-CDF.

Yanyi Chu is a Ph.D. candidate at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University. She works on drug discovery through machine learning methods.
Aman Chandra Kaushik is an assistant professor at the School of Medicine, Jiangnan University, Wuxi, China. His main research interests focus on machine learning algorithms and their applications in precision based medicines.
Xiangeng Wang is a master student at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University. He works on multi-label learning and interpretable machine learning with the application of biomedicine.
Wei Wang is a Ph.D. candidate at the School of Mathematical Sciences, Shanghai Jiao Tong University. He works on statistical learning algorithms for drug discovery.
Yufang Zhang is a master student at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University. She works on drug-target interactions prediction by using machine learning methods.
Xiaoqi Shan is a master student at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University working in the study of drug metabolism, especially the multi-label classification models for the prediction of CYP450 enzyme-substrate selectivity.
Dennis Russell Salahub is a full professor at the Department of Chemistry, University of Calgary, Fellow Royal Society of Canada, and Fellow of the American Association for the Advancement of Science.
Yi Xiong is an associate professor at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University. His main research interests focus on machine learning algorithms and their applications in the protein sequence-structure-function relationship and biomedicine.
Dong-Qing Wei is a full professor at the School of Life Sciences and Biotechnology, Shanghai Jiao Tong University. His main research areas include structural bioinformatics and biomedicine.
Submitted: 16 June 2019; **Received (in revised form):** 1 November 2019

Key words: drug-target interaction; machine learning; ensemble learning; cascade deep forest

Introduction

Drug discovery is the process of identifying new candidate compounds with potential therapeutic effects, during which the prediction of drug-target interactions (DTIs) is an essential step. Drugs play a significant role in the human body by interacting with various targets. Proteins represent an important type of target, and their functions can be enhanced or inhibited by drugs to achieve phenotypic effects for therapeutic purposes [1]. However, the number of approved drugs is relatively small, mainly due to the possible adverse effects of the multi-targeting of drugs. Currently, a large number of researches have focused on DTI prediction because it is an essential tool in the context of drug repurposing. Since experimental determination of DTIs is both time- and resource-consuming, the development of efficient computation methods is highly desirable to make full use of the heterogeneous biological data of known DTIs to understand the mechanism of action of drugs in the human body.

Over the past decades, a substantial number of computational methods have been developed for the prediction of DTIs. As suggested by a series of recent review articles [2–12], the brief categorization of the available DTI prediction methods is summarized in Figure 1. Early attempts in traditional approaches for computational prediction of DTIs include the ligand-based [13] and target-based [14] approaches. The ligand-based methods compare a query ligand to a set of known ligands with target proteins. The prediction results of ligand-based methods may become unreliable in cases when the number of known ligands with target proteins is insufficient. The target-based methods such as docking methods consider the 3D structures of target proteins. However, these approaches are extensively time-consuming and even cannot work when the structural information is not available for some targets such as membrane proteins. Moreover, dealing with the flexibility of a target protein can be very challenging. Therefore, it is difficult to use the target-based approaches on a genome-wide scale. To overcome the limitations of traditional approaches and conduct large-scale DTI prediction, the chemogenomic methods have attracted much interest with the accessibility of big data sources such as genome, phenome, drug chemical structures, biological interactome, and biological bioassays, which provide a useful way to extract different information from the drug side (chemical space) and target side (genomic feature space) simultaneously to predict novel DTIs.

Following the way to formulate the DTI prediction problem, the chemogenomic methods can be divided into two major classes of solving strategies: network/graph-based methods [15–47] and machine learning-based methods [48–88]. In network/graph-based methods, the interaction space of drugs and targets is represented as a bipartite graph, in which the nodes are drugs/targets and the edges are the interactions between drugs and targets. Therefore, the graph-based and network-based analysis methods can be applied in the task of predicting novel DTIs to infer the missing links in the graph/network. Similarly, the bipartite graph can be transformed into an association matrix. The hidden associations can be inferred by methods such as matrix factorization. Zhang *et al.* [6] summarized the recent advances of network-based models in predicting DTIs. In machine learning-based methods, the problem of DTI prediction is formulated as a binary classification task to predict whether a drug-target pair is DTI or not. On one hand, the information about drugs and targets are represented as features, and the interactions between drugs and targets are denoted as class labels. On the other hand, the interaction network inference problem can be transformed into a binary classification task between drug-target pairs using pairwise kernel functions. The recent machine learning-based methods are composed of semi-supervised models and supervised models. In semi-supervised machine learning-based methods, they utilize both a small number of available labeled samples (known DTIs in the data set) and a large number of unlabeled samples (all the unknown DTIs in the data set). A few of semi-supervised models were developed for DTI prediction, such as NetLapRLS [49], NetCBP [87], and ILRLS [88]. Moreover, there are a few other interesting methods, such as the text mining-based method [89] and a two-layer graphical model (called restricted Boltzmann machine) [90]. More recently, a number of deep learning-based methods have been developed for DTI prediction [91–95].

Motivated by the previous studies [96, 97], we develop a cascade deep forest (CDF)-based model to further improve the performance of predicting DTIs. In the proposed method, we firstly utilize path-category-based multi-similarities features (named PathCS) based on the heterogeneous graph of DTIs. Then, we apply the CDF model under three experimental settings through five repeated 10-fold cross-validation (CV) in four representative data sets, and the performance evaluation is conducted by using the AUPR, AUC, and F_2-score metrics and the average among them. Furthermore, the statistical hypothesis test is used to evaluate the statistical significance of the results. Finally, we verify that the proposed DTI-CDF method is significantly better than the traditional ensemble learning-based approaches such as random forest (RF) and XGBoost (XGB), deep learning-based approaches such as the deep neural network (DNN), and the state-of-the-art methods available (i.e., DDR [79]). More importantly, our method predicted 1352 new DTIs which have been supported by KEGG and DrugBank databases.

Materials and methods

Data sets

In our work, the four data sets compiled by Yamanishi *et al.* [15] were used as a benchmark to evaluate the performance of the proposed DTI-CDF method in DTI prediction. The four data sets are separated and named by the type of target proteins of the drugs: enzymes (E), ion channels (IC), G-protein-coupled receptors (GPCR) and nuclear receptors (NR). All drugs in these data sets are approved drugs that were searchable at the KEGG DRUG database [98] which is a comprehensive drug information resource for drugs approved in Japan, the United States, and Europe. These data sets contain known human DTIs retrieved from the KEGG BRITE [98], BRENDA [99], SuperTarget [100], and DrugBank [101] databases. These resources are highly reliable so that the results obtained by the use of these data sets have high reliability. Therefore, it is generally considered as the gold-standard data sets.

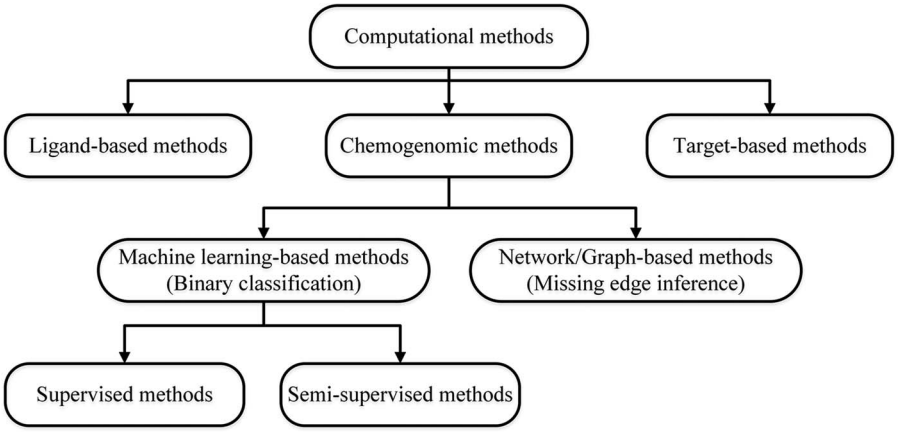

Figure 1. The categories of computational methods for DTI prediction.

Table 1. Summary of samples on the four data sets

Data sets	Known interactions	Unknown interactions	Drugs	Targets
NR	90 (6.41%)	1314 (93.59%)	54	26
GPCR	635 (3%)	20 550 (97%)	223	95
IC	1476 (3.45%)	41 364 (96.55%)	210	204
E	2926 (1%)	292 554 (99%)	445	664

In order to simulate more practically, we consider the entire space of the DTIs in these four data sets. The known DTIs are considered as positive samples, and the negative data contains all unknown or non-existing DTIs. It is worth noting that the number of positive samples is much lower than the number of negative samples. Thus, these four data sets are severely unbalanced, as shown in Table 1.

Feature construction

PathCS [79] is a hybrid feature based on the heterogeneous weighted graph of DTIs, containing drugs, targets, and their similarities or interactions. In this graph, the edge between two target nodes or two drug nodes represents their similarities, and the weight of the edge is the similarity value between two linked nodes. The edge between a target and a drug denotes a known DTI, and the weight is equal to 1.

There are six types of kernels used in this study to generate similarity profiles for drugs and targets, defined as follows:

(i) Protein kernels. We use amino acid sequences of the proteins to generate the spectrum kernel [102] and set the subsequence length k as 4.

(ii) Drug kernels. There are three side-effects kernels as drug information sources. The first resource is obtained from the SIDER [103] database, which contains information on marketed drugs and their adverse reactions. For each side-effect classification, a binary (absence or presence) profile was used to represent drugs. The other two pharmacological profiles are derived from the Food and Drug Administration's adverse event reporting system [104] on the basis of frequency and binary information of side-effect classifications, respectively. These three pharmacological profiles are used to generate similarity profiles through the weighted cosine correlation coefficient. And if a drug is not in the data resources, its similarity is assigned as 0.

(iii) Gaussian interaction profile (GIP) kernels. Firstly, the interaction profile of a drug is a binary vector based on the known DTI network, in which the absence or presence of interaction with every target in the network is assigned as 0 or 1 [50]. A similar definition fits for the interaction profile of a target. In the target interaction profile for drugs, the i-th column represents the target interaction profile y_{d_i} of the drug d_i. Furthermore, the GIP similarity between a pair of drugs d_i and d_j can be computed between the two corresponding columns of the target interaction profile:

$$K_{\mathrm{GIP}}(d_i, d_j) = e^{-\gamma_d \left\| y_{d_i} - y_{d_j} \right\|^2} \tag{1}$$

where the parameter γ_d controls the kernel bandwidth given by

$$\gamma_d = \frac{1}{n_d} \sum_{i=1}^{n_d} \left| y_{d_i} \right|^2 \tag{2}$$

where n_d is the number of drugs. This kernel is independent of the size of the data set because of normalization.

The GIP similarity between a pair of targets is calculated in a similar way. However, the GIP kernel cannot be computed for a new drug (or a new target), which does not have any targets (or drugs) to interact with in the training data set. For the calculation of this kernel, we adopted the method of neighbor-based interaction-profile inferring [23]. The inference of the similarity of an unknown drug (or target) to a particular target (or drug) is done using the five neighbors of the unknown drug (or target), expressed as the ratio of the sum of the similarities of five neighbors that interact with the particular target (or drug) to the sum of similarities of all neighbors.

After obtaining the above similarity measures, the first step is to combine the multiple similarity measures of drugs (or targets) into one fused matrix [105] to build a heterogeneous DTIs graph

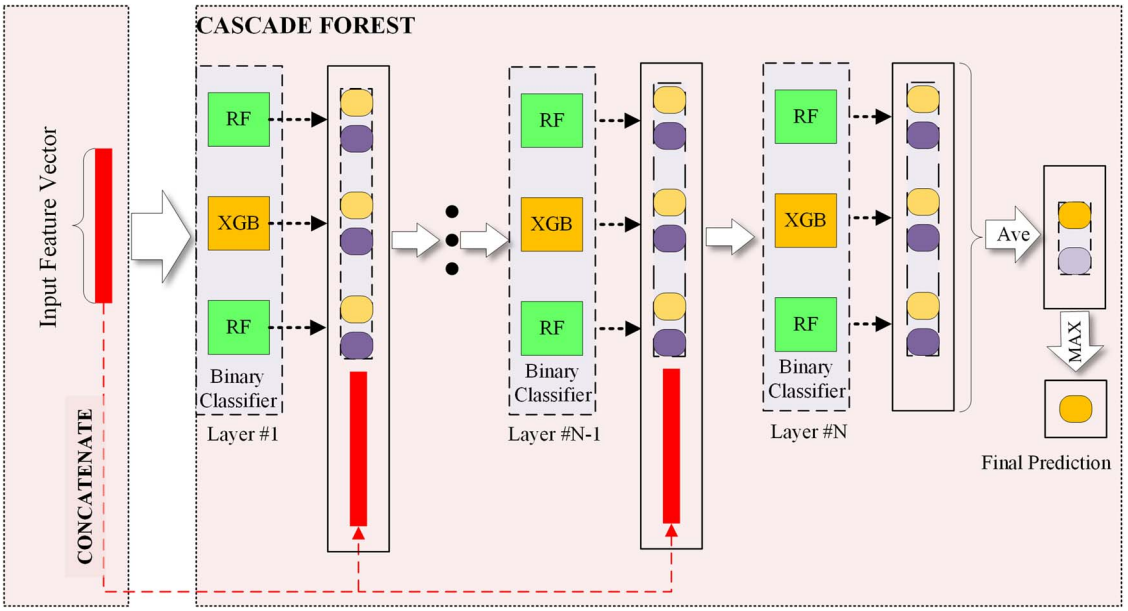

Figure 2. This machine learning model is composed of an input feature vector, a CDF classifier, and a final prediction. In particular, CDF is the core unit of the model, which has six variants in this study. In each variant, each layer consists of a different number of RF and XGB binary classifiers, and different layers own the same structure. The figure shows one special model in which each layer has two RF learners and one XGB learner, denoted as RF2-XGB1. Other variants are RF2, XGB2, RF1-XGB2, RF1-XGB1, and RF2-XGB2, respectively.

and then extract PathCS for each drug-target pair. The path category is defined by a path structure that starts at a drug node and ends up at a target node such as to set the path length to 2 or 3. Path categories are as follows: drug-drug-target, drug-target-target, drug-drug-drug-target, drug-drug-target-target, drug-target-drug-target, and drug-target-target-target. We define two normalized matrices N_1^h and N_2^h according to the above six path categories C^h, $h = 1, 2, \cdots, 6$. For a specific drug d_i and a specific target t_j, we denote one path from d_i to t_j as p_q and the set of paths is R_{ijh}. In addition, the path between d_i and t_j is built by the intermediate nodes which are restricted to be the five nearest neighbors of d_i and t_j, respectively. Thus, the N_1^h and N_2^h with elements $n_1^h(i,j)$ and $n_2^h(i,j)$, respectively, are computed as follows:

$$n_1^h(i,j) = \frac{\sum_{\forall q: p_q \in R_{ijh}} \prod_{\forall w_x \in p_q, p_q \in R_{ijh}} w_x}{\sum_j \sum_{\forall q: p_q \in R_{ijh}} \prod_{\forall w_x \in p_q, p_q \in R_{ijh}} w_x} \quad (3)$$

$$n_2^h(i,j) = \frac{\max_{\forall q: p_q \in R_{ijh}} \prod_{\forall w_x \in p_q, p_q \in R_{ijh}} w_x}{\sum_j \max_{\forall q: p_q \in R_{ijh}} \prod_{\forall w_x \in p_q, p_q \in R_{ijh}} w_x} \quad (4)$$

Classification algorithm

Firstly, we generate PathCS as the input feature vector for each DTI. Secondly, a CDF classifier [106] is used to predict DTIs. In this process, the new category probability vector in the previous layer and the original input feature vector are used as the next layer input, and the final category probability vector is the output through multiple learners. When building a CDF model (Figure 2), it is important to determine the machine learner used for each layer. In the model, we set the number of learners of each layer from 2 to 6, and RF [107], which has presented satisfactory performance in another classification task [108], and XGB [109] are used as learners to follow the "good but distinguishable" principle. In addition, the depth of layers is determined by the trend of evaluation metrics.

Table 2. Summary of the corresponding DTIs information in the test data of three experimental settings

Experimental settings	Drugs	Targets	Interactions
S_P	Known	Known	New
S_D	New	Known	New
S_T	Known	New	New

Experimental settings

In this study, we evaluate three experimental settings as Table 2 shows, which include most of the conditions for DTI prediction. For these experimental settings, S_P, S_D, and S_T represent the corresponding DTI values of certain drug-target pairs, drugs, targets in the training set are missed but existed in the test sets. In Table 2, the subjects which are new indicate that no corresponding subjects exist in the training data.

Performance evaluation

In order to facilitate the comparison with other methods, we followed previous studies [31, 79] as the benchmark and conducted the 10-fold CV test for each experimental setting of each data set, and the above process was repeated 5 times using different random seeds. It is worth noting that the CV used in this study is different from the traditional CV, i.e., the performance of the test set is only used to evaluate the model performance but not for model selection, which is like using a holdout test in each experiment of the CV.

For each fold of each predictive model, the following metrics are calculated:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Table 3. The hypothesis test results

Model/method	Metrics	n	$t_{0.05}\ (n-1)$	t_0	p-value	δ
CDF versus DNN	AUPR, AUC, F_2-score	36	1.69	6.41	1.11E − 7	1.07
CDF versus RF	AUPR, AUC, F_2-score	36	1.69	6.56	7.02E − 8	1.09
CDF versus XGB	AUPR, AUC, F_2-score	36	1.69	7.05	1.60E − 8	1.18
DTI-CDF versus DDR	AUPR, AUC	24	1.71	6.08	1.68E − 6	1.24

The DTI-CDF and DDR are the proposed method in this study and the state-of-the-art method, respectively.

$$\text{True positive rate} = \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{6}$$

$$\text{False positive rate} = \frac{\text{FP}}{\text{FP}+\text{TN}} \tag{7}$$

$$F_\beta - \text{score} = \frac{(1+\beta^2)\cdot \text{Precision}\cdot \text{recall}}{\beta^2\cdot \text{Precision} + \text{recall}} \tag{8}$$

where TP is true positive, FP is false positive, FN is false negative, and TN is true negative. We plot the precision-recall curve (PR curve) based on different precision and recall, and the receiver operating characteristic curve (ROC curve) based on different recall and false-positive rate, respectively, under the condition of different classified cutoff values. We define AUPR and AUC as the area under the PR curve and the ROC curve, respectively. Since the positive samples and negative samples in each data set are highly unbalanced, the AUPR provides a better performance estimate relative to AUC because it more severely penalizes the false positives. On the other hand, AUC avoids the subjectivity of threshold selection, as does AUPR. Therefore, we introduce $F_\beta - \text{score}$, and set β to 2, called F_2-score, to increase the effect of recall on this metric because small FN may reduce the possibility that new DTIs cannot be identified. For each experiment setting of each data set, the AUPR, AUC, and F_2-score are calculated as a measure of model performance as follows:

$$\text{AUPR} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{10}\text{AUPR}_{i,j}}{50} \tag{9}$$

$$\text{AUC} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{10}\text{AUC}_{i,j}}{50}, \tag{10}$$

$$F_2 - \text{score} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{10}F_2 - \text{score}_{i,j}}{50} \tag{11}$$

where i represents the i-th repeated trials and j represents the j-th fold of CV. In addition, the average of the above three metrics can be calculated as a weighted performance metric [110].

Statistical hypothesis test

The statistical hypothesis test is used in this study to further explore the statistical significance of the difference between the proposed method and the other method. Differences in the results of different prediction methods are caused by a variety of factors, such as data composition, training model, and experimental setting. In order to exclude other factors and only consider the differences caused by the point we considered, the one-sided paired t-test that is a pairwise comparison method based on paired data is employed. Firstly, the difference $d_i \in \mathbf{D}$ of performance metrics, such as AUPR, AUC, and F_2-score, based on 12 experimental conditions (i.e., four data sets under three experimental settings) between the two methods are calculated. It is assumed that the difference d_i are all from the normal distribution $N(\mu_d, \sigma^2)$, where both μ_d and σ^2 are unknown. Then, a statistical hypothesis test is performed based on the data

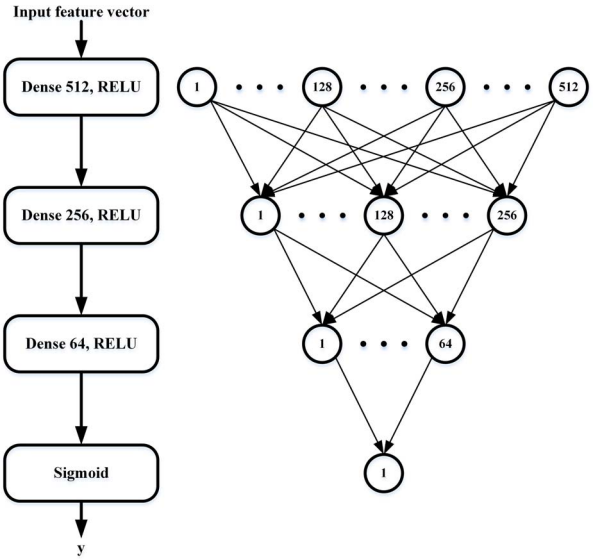

Figure 3. The structure of the DNN used in this study. The Dense represents the fully connected layer, RELU and Sigmoid are activate functions.

obtained above. If the two methods are no different on performance, the difference d_i between each pair of data belongs to a random error, and the random error can be considered to obey a normal distribution with a mean of zero. Assuming that there is no difference between the above two methods, the test hypothesis is as follows:

$$H_0 : \mu_d = 0, H_1 : \mu_d > 0 + \Delta \tag{12}$$

By the t-test of a single population means using the normal distribution, the rejection domain is:

$$t = \frac{\overline{D}}{s/\sqrt{n}} \geq t_\alpha\ (n-1) \tag{13}$$

where $\overline{D}$ is the mean of the sample, s is the SD of the sample, n is the sample size, α is the significance level, and Δ is equivalent to the effect size of mean difference, defined as $\Delta = u_d/\sigma$. In order to ensure that only when the proposed method is far superior to another, it can be tested with a high probability $1-\beta$; we set $\alpha = 0.05$, $\Delta = 0.9$, and $\beta = 0.01$. Under these conditions, a sample size n not less than 21 is required. The rejection domain and the actual effect size of mean difference are

$$t = \frac{\overline{D}}{s/\sqrt{n}} \geq t_{0.05}\ (n-1) \tag{14}$$

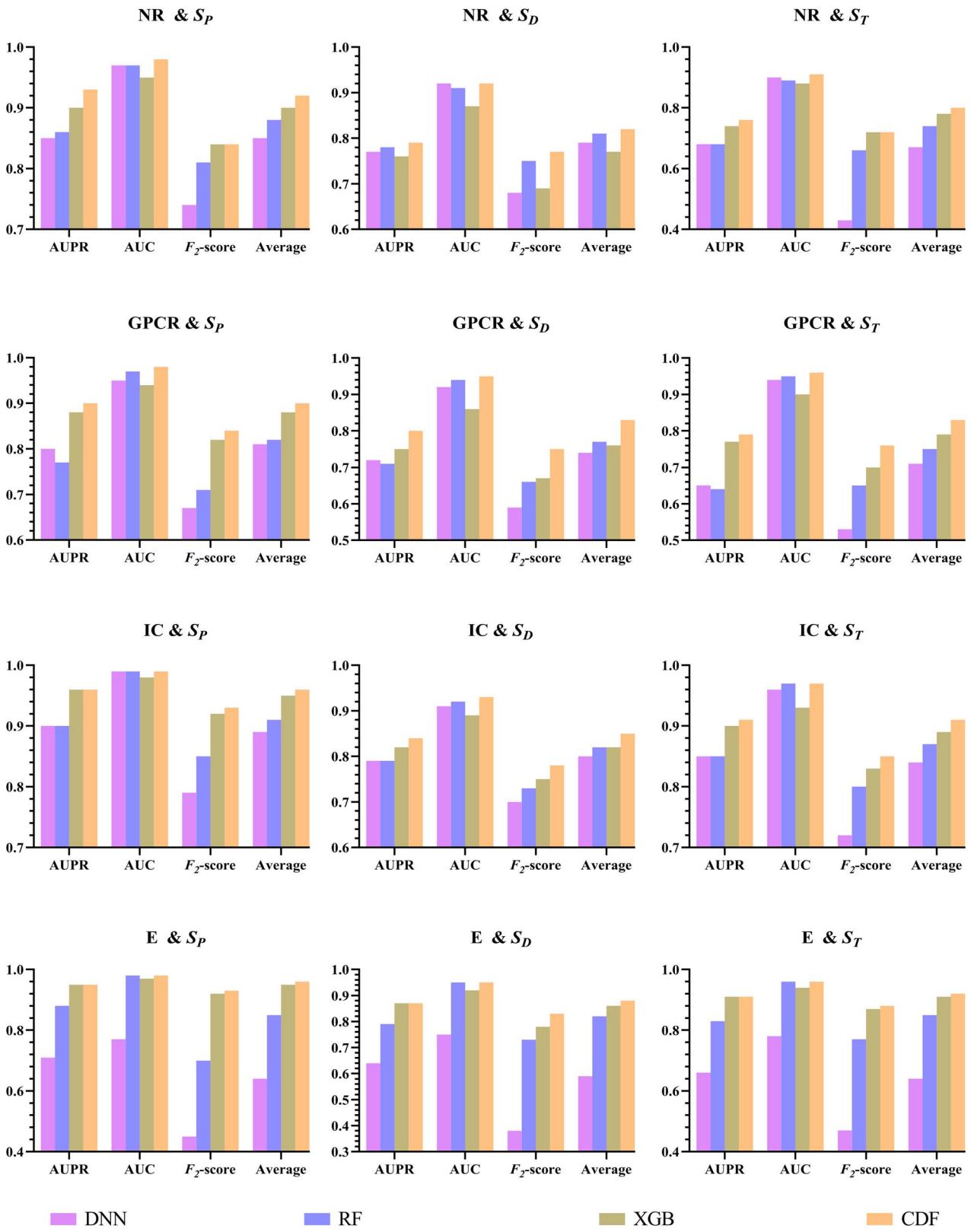

Figure 4. The comparison of DNN, RF, XGB, and CDF in four data sets such as NR, GPCR, IC, E under three experimental settings (i.e., S_P, S_D and S_T). The evaluation metrics are AUPR, AUC, F_2-score and the average of them.

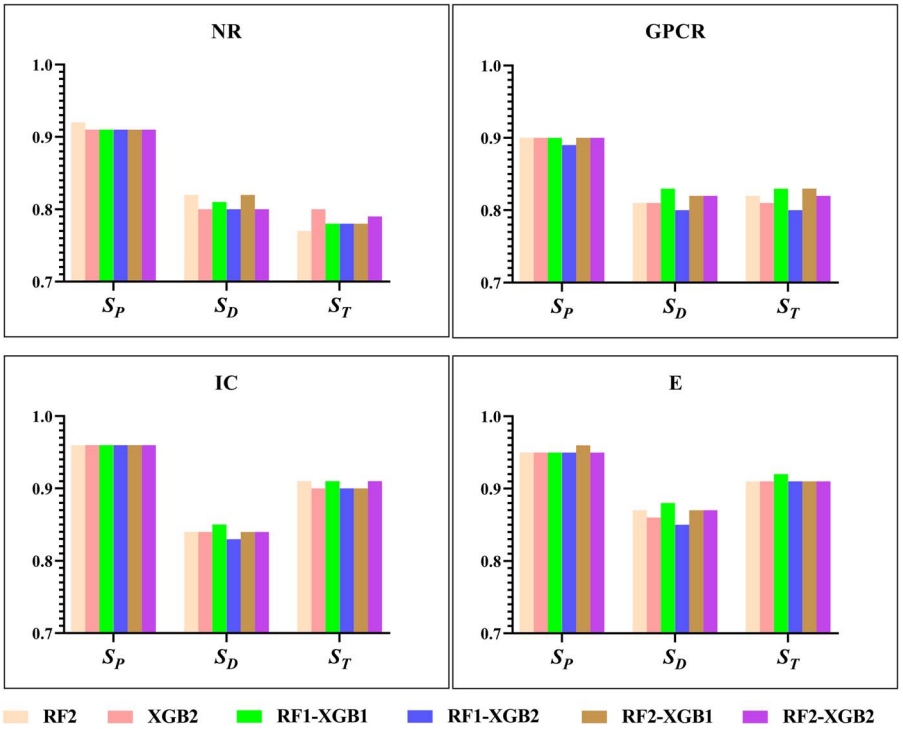

Figure 5. The performance comparison of six combinations of RF and XGB in a layer of the CDF model in four data sets such as NR, GPCR, IC, E under three experimental settings (i.e., S_P, S_D and S_T), including two RFs (RF2), two XGBs (XGB2), one RF and one XGB (RF1-XGB1), one RF and two XGBs (RF1-XGB2), two RFs and one XGB (RF2-XGB1), and two RFs and two XGBs (RF2-XGB2). The value represents the average of AUPR, AUC, and F_2-score.

$$\delta = \frac{\overline{D}}{s} \tag{15}$$

Substituting d_i into the above formula yields the observed value t_0 of t; then the *p*-value of the right-tailed *t*-test can be calculated by

$$p = P\{t \geq t_0\} \tag{16}$$

Results and discussion

Comparison of CDF model with the deep learning model

Recently, DNN or deep learning has achieved great success in many areas including bioinformatics [111]. However, it still has apparent deficiencies. It is well known that the training of DNN usually requires a large amount of data, so its implementation on tasks with small-scale data is difficult. Although we are in the era of big data, many practical tasks still lack a sufficient amount of labeled data due to high labeling cost, resulting in poor performance of DNN in these tasks. Secondly, DNN is a very complex model, and the training processes often require powerful computing devices. More importantly, DNN has too many hyper-parameters, and the learning performance is heavily dependent on their careful tuning that makes the training very difficult. In addition, the theoretical analysis of DNN is extremely difficult because too many interference factors are combined with almost unlimited parameter configurations. As a large amount of training data is used in DNN, and the learning ability of the model must be large, we can conclude that DNNs are more complicated than ordinary learning models.

In this study, we developed a CDF model [106] with which it is possible to achieve performance that does not have the above drawbacks and can compete with DNN. It is a deep ensemble framework that cascades traditional machine learning models (such as RF and XGB). Compared to DNN, the CDF model has fewer hyper-parameters, and it is easier to train. In addition, unlike most types of DNN with fixed model complexity, the CDF model can stop the increase of the number of layers by terminating the training properly, and the complexity of the model can be adaptively scaled, making the CDF model not limited to large-scale training data but also on small-scale training data. Moreover, if a tree-based approach is chosen as the base-learner, CDF will be easier to theoretically analyze than DNN.

To clarify the superiority of the CDF model compared with DNN whose structure is shown in Figure 3, we compared them in this study, and the results are shown in Figure 4. It has been shown in our experiments that CDF achieved highly competitive performance in comparison to DNN since all results in different experimental conditions are better than that of DNN. The reason may be that the sample sizes of the four data sets used in this study ranged from 10^3 to 10^5. In addition, the number of positive and negative samples in each data set is highly unbalanced; too few positive samples make DNN which is based on a large amount of training data unable to exert its advantages. Moreover, the feature dimension used in this study is low, and deep learning has great advantages in the representation learning of ultrahigh dimensional data.

Furthermore, we use the one-sided paired *t*-test, and the test results (such as *p*-value) are listed in Table 3. It is shown that the performance of CDF is significantly better than that of DNN on this task.

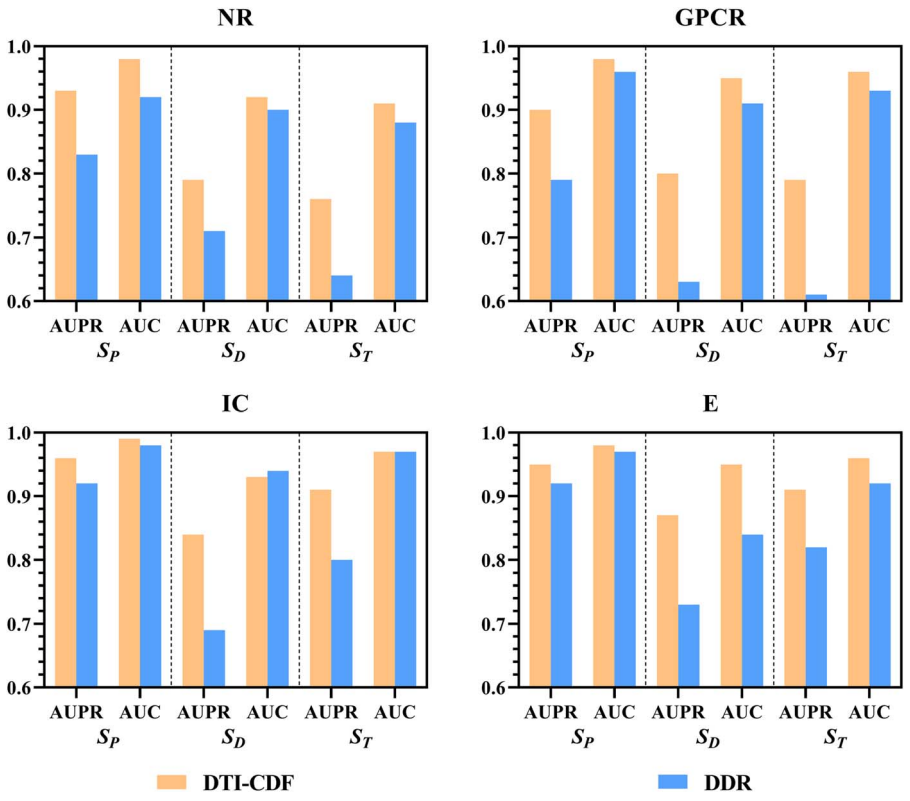

Figure 6. The comparison of the proposed DTI-CDF method with the state-of-the-art DDR method in four data sets such as NR, GPCR, IC, E under three experimental settings (i.e., S_P, S_D and S_T).

Comparison of CDF model with the traditional ensemble learning model

One of the main types of classification methods in machine learning is ensemble learning which completes learning tasks by constructing and combining multiple learners. The ensemble learning is usually possible to obtain generalization performance superior to that of a single learner and to solve the imbalanced classification problem to some extent. According to the generation method of individual learners, the current ensemble learning methods can be divided into two categories: boosting and bagging, which are focused on reducing bias and variance, respectively. In order to construct a CDF model with low variance and low deviation, this study combines boosting (i.e., XGB) and bagging (i.e., RF) where XGB and RF are chosen as the base-classifiers.

XGB is a scalable tree boosting method that adds regular terms to the cost function, which can control the complexity of the model and prevent overfitting. At the same time, it uses a second-order Taylor expansion approximation to the cost function, which makes the approximation of the objective function closer to the actual value, thus improving the prediction accuracy. In addition, the calculations of XGB can speed up by using some techniques, such as column subsampling [109].

On the basis of bagging, RF introduces the random feature selection in the training process of the decision tree. That is, the diversity of the RF not only comes from the sample disturbance but also from the feature disturbance, and this improves the generalization performance. Moreover, RF is easy to implement and exhibit powerful performance in many real-world tasks, especially in bioinformatics.

In this study, we use these two base-learners as the benchmark models in order to compare the CDF model with the traditional ensemble learning models. The results (Figure 4) indicate that the CDF model is superior to RF and XGB under all experimental conditions. The reason is that the diversity and complementary of base-learners of the CDF model improve the classification performance. Moreover, the multilayer characteristics of CDF are more fully exploited for features.

In order to obtain the significance level between the CDF and RF model as well as the CDF and XGB model, we also carried out the hypothesis test. For the above two results (Table 3), we can reasonably believe that the CDF model performed significantly better than RF and XGB.

It is worth noting that in order to reduce the complexity of the model, only RF and XGB are considered as base-learners in this study. In the future, all possible individual learners can be enumerated, and individual learners with the best classification effect can be selected for the ensemble.

Effect of the combination of base-learners in CDF model

In the previous study, we have demonstrated the optimal performance under the test set by using one of the architectures of CDF. Moreover, in order to clarify the robustness of our model, we first validated the other different architectures of CDF on the NR data set and found that the prediction performance is approximately equal to the optimal one we have discussed as shown in Figure 5. Thus, it demonstrates the robustness of our CDF model against the model selection, which will avoid the large amount of work in the parameter tuning. More significantly, we further perform

Table 4. Summary of the predicted and new reported DTIs of the four data sets, including the number, the number that has been reported by the KEGG and DrugBank databases

Data sets	Predicted DTIs	New reported DTIs		
		KEGG	DrugBank	Total
NR	84	55	72	74
GPCR	500	229	303	338
IC	1158	322	284	460
E	1650	299	326	480

the experiment on other data sets. It is found that the prediction accuracy is approximated to be equal with each other as Figure 5 shows. It indicates that our CDF model preserves the universality over different data sets such that it can be migrated to other DTI applications. It indicates that the CDF model is not very sensitive to parameter settings. Thus, we do not need to conduct the large-scale parameter tuning including the selection of the optimal combinations of base-classifiers, which is also one of the advantages when compared to DNN.

Comparisons with the state-of-the-art algorithm (DTI-CDF versus DDR)

For the four data sets in this work, the DDR method is proved as the most powerful approach in the prediction of DTIs under the same experimental conditions (i.e., five repeated trails of 10-fold CV under three experimental settings of each data set). Therefore, in the present study, we only compare our method DTI-CDF with DDR. The experimental results demonstrate that DTI-CDF achieves better performance than DDR under the same conditions (Figure 6).

In order to validate the significant difference between DTI-CDF and DDR, we also carried out the one-sided paired *t*-test. The test results are listed in Table 3. It is shown that the DTI-CDF method is significantly better than the DDR method.

Predicted DTIs reported in KEGG and DRUGBANK databases

In order to evaluate the utility of the DTI-CDF method, it is possible to effectively predict DTIs that are real but not yet contained in the data sets of this study (called new DTIs). We refer to the KEGG and DrugBank databases, and the DTIs predicted by the DTI-CDF method on the four data sets are searched, confirming that 1352 new DTIs are supported by the reference databases (Table 4), and the information of these new DTIs is publicly available at https://github.com/a96123155/DTI-CDF/tree/master/3_new_DTIs. This reflects the credibility of the DTI-CDF method, and other DTIs that have not yet been reported but are predicted by the method are likely to be real.

Conclusions

Identification of DTIs is fundamental to both new drug discovery and new uses for existing drugs. In the present study, we propose a DTI-CDF method to predict DTIs, which utilizes similarity information for drugs and targets as the input of our algorithm for DTIs prediction. We use AUPR, AUC, F_2-score, and their average to evaluate the performance of the DTI-CDF method under three different experimental settings based on gold-standard data sets, and almost all of them are superior to the current top-performing method DDR. It is further proved that the performance of the DTI-CDF method is significantly better than other existing methods when a known DTI is missing from the training data, especially in searching targets for new drugs (S_D setting) and finding drugs for new targets (S_T setting). Experimental results further demonstrated that the DTI-CDF method presents higher predictive performance than the deep learning-based method, such as DNN, and traditional ensemble learning models such as RF and XGB. Moreover, 1352 predicted new DTIs are proved to be true cases by KEGG and DrugBank databases.

More recently, various types of noncoding RNAs (ncRNAs) have been identified. Increasing evidence has shown that these ncRNAs may affect gene expression and disease progression, making them a new class of targets for drug discovery. It thus becomes important to understand the relationship between ncRNAs and drug targets and further identify the association between small molecules and ncRNAs [112–116]. On one hand, we will extend the data sets to include the new type of targets (i.e., ncRNAs) and work on the prediction of drug-ncRNA interactions. On the other hand, most of the computational methods have focused on the binary classification problem to predict whether a drug-target pair interacts or not. However, there are very few computational approaches that can predict the drug-target affinity [117, 118], which will be explored in our future work.

Key Points

- Prediction of DTIs is very important in drug discovery, especially using computational methods such as machine learning methods. A dominant issue in the prediction of DTIs is the absence of a list of true negative samples.
- The proposed DTI-CDF method extracts features from the heterogeneous DTI weighted graph as an input feature vector of the CDF-based model.
- DTI-CDF method uses more than one source of information at a time. Different aspects of drugs and targets are represented by these sources and are used simultaneously to improve prediction performance.
- CDF is a courageous attempt in the field of predicting DTIs, which enables the deeperization of traditional machine learning models and better performance than traditional ensemble learning and deep learning methods.
- DTI-CDF method achieves the best performance in different data sets and prediction settings, indicating the robustness of the method.
- There are 1352 predicted new DTIs that have been supported by KEGG and DrugBank databases, which indicates the usefulness of the proposed DTI-CDF method.

Acknowledgments

The authors would like to thank Prof. Caihong Yang for the technical assistance in mathematics, Tailong Xiao for the fruitful discussion of the article, and everyone in the laboratory of Prof. Dong-Qing Wei for providing help on the using and testing of the source code of the model.

Funding

National Key Research and Development Program of China (Contract No. 2016YFA0501703) and National Natural Science Foundation of China (Grant No. 31601074, 61872094, 61832019).

References

1. Santos R, Ursu O, Gaulton A, *et al*. A comprehensive map of molecular drug targets, nature reviews. *Drug Des Discov* 2017;**16**:19–34.
2. Kuhn M, Campillos M, Gonzalez P, *et al*. Large-scale prediction of drug-target relationships. *FEBS Lett* 2008;**582**:1283–90.
3. 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**:734–47.
4. Chen X, Yan CC, Zhang X, *et al*. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**:696–712.
5. Cheng T, Hao M, Takeda T, *et al*. Large-scale prediction of drug-target interaction: a data-centric review. *AAPS J* 2017;**19**:1264–75.
6. Zhang SW, Yan XY. Some remarks on prediction of drug-target interaction with network models. *Curr Top Med Chem* 2017;**17**:2456–68.
7. Anusuya S, Kesherwani M, Priya KV, *et al*. Drug-target interactions: prediction methods and applications. *Curr Protein Pept Sci* 2018;**19**:537–61.
8. Chen R, Liu X, Jin S, *et al*. Machine learning for drug-target interaction prediction. *Molecules* 2018;**23**.
9. Ezzat A, Wu M, Li XL, *et al*. Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform* 2019;**20**:1337–1357.
10. Ding Y, Tang J, Guo F. The computational models of drug-target interaction prediction. *Protein Pept Lett* 2019.
11. Zhang W, Lin W, Zhang D, *et al*. Recent advances in the machine learning-based drug-target interaction prediction. *Curr Drug Metab* 2019;**20**:194–202.
12. Zhao Q, Yu H, Ji M, *et al*. Computational model development of drug-target interaction prediction: a review. *Curr Protein Pept Sci* 2019;**20**:492–4.
13. Keiser MJ, Roth BL, Armbruster BN, *et al*. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**:197–206.
14. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. *Proteins* 2006;**65**:15–26.
15. 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**:i232–40.
16. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 2009;**25**:2397–403.
17. 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**:i246–54.
18. Zhao S, Li S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS One* 2010;**5**:e11764.
19. Chen X, Liu MX, Yan GY. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol BioSyst* 2012;**8**:1970–8.
20. 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**:e1002503.
21. Alaimo S, Pulvirenti A, Giugno R, *et al*. Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics* 2013;**29**:2004–8.
22. Alaimo S, Bonnici V, Cancemi D, *et al*. DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst Biol* 2015;**9**(Suppl 3):S4.
23. Mei JP, Kwoh CK, Yang P, *et al*. Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics* 2013;**29**:238–45.
24. Wang W, Yang S, Li J. Drug target predictions based on heterogeneous graph inference. *Pac Symp Biocomput* 2013;53–64.
25. Kim S, Jin D, Lee H. Predicting drug-target interactions using drug-drug interactions. *PLoS One* 2013;**8**:e80129.
26. Seal A, Ahn YY, Wild DJ. Optimizing drug-target interaction prediction based on random walk on heterogeneous networks. *J Chem* 2015;**7**:40.
27. Yan XY, Zhang SW, Zhang SY. Prediction of drug-target interaction by label propagation with mutual interaction information derived from heterogeneous network. *Mol BioSyst* 2016;**12**:520–31.
28. Ba-Alawi W, Soufan O, Essack M, *et al*. DASPfind: new efficient method to predict drug-target interactions. *J Chem* 2016;**8**:15.
29. Lan W, Wang JX, Li M, *et al*. Predicting drug-target interaction using positive-unlabeled learning. *Neurocomputing* 2016;**206**:50–7.
30. Liu Y, Wu M, Miao C, *et al*. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput Biol* 2016;**12**:e1004760.
31. Nascimento AC, Prudencio RB, Costa IG. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinform* 2016;**17**:46.
32. Bolgar B, Antal P. VB-MK-LMF: fusion of drugs, targets and interactions using variational Bayesian multiple kernel logistic matrix factorization. *BMC Bioinform* 2017;**18**:440.
33. Durán C, Daminelli S, Thomas JM, *et al*. Pioneering topological methods for network-based drug-target prediction by exploiting a brain-network self-organization theory. *Brief Bioinform* 2018;1183–202.
34. Ezzat A, Zhao P, Wu M, *et al*. Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans Comput Biol Bioinform* 2017;**14**:646–56.
35. Hao M, Bryant SH, Wang Y. Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Sci Rep* 2017;**7**:40376.
36. Lu Y, Guo Y, Korhonen A. Link prediction in drug-target interactions network using similarity indices. *BMC Bioinform* 2017;**18**:39.
37. Luo Y, Zhao X, Zhou J, *et al*. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**:573.
38. Peska L, Buza K, Koller J. Drug-target interaction prediction: a Bayesian ranking approach. *Comput Methods Prog Biomed* 2017;**152**:15–21.

39. Wu Z, Cheng F, Li J, *et al*. SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief Bioinform* 2017;**18**:333–47.
40. Zhang W, Chen Y, Li D. Drug-target interaction prediction through label propagation with linear neighborhood information. *Molecules* 2017;**22**:2056.
41. Zhang X, Li L, Ng MK, *et al*. Drug-target interaction prediction by integrating multiview network data. *Comput Biol Chem* 2017;**69**:185–93.
42. Zong N, Kim H, Ngo V, *et al*. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics* 2017;**33**:2337–44.
43. Lee I, Nam H. Identification of drug-target interaction by a random walk with restart method on an interactome network. *BMC Bioinform* 2018;**19**:208.
44. Seal A, Wild DJ. Netpredictor: R and shiny package to perform drug-target network analysis and prediction of missing links. *BMC Bioinform* 2018;**19**:265.
45. Wang M, Tang C, Chen J. Drug-target interaction prediction via dual Laplacian graph regularized matrix completion. *Biomed Res Int* 2018;**2018**:1425608.
46. Ban T, Ohue M, Akiyama Y. NRLMFbeta: beta-distribution-rescored neighborhood regularized logistic matrix factorization for improving the performance of drug-target interaction prediction. *Biochem Biophys Rep* 2019;**18**:100615.
47. Yan XY, Zhang SW, He CR. Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. *Comput Biol Chem* 2019;**78**:460–7.
48. Jacob L, Vert JP. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics* 2008;**24**:2149–56.
49. Xia Z, Wu LY, Zhou X, *et al*. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst Biol* 2010;**4**(Suppl 2):S6.
50. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 2011;**27**:3036–43.
51. Yu W, Jiang Z, Wang J, *et al*. Using feature selection technique for drug-target interaction networks prediction. *Curr Med Chem* 2011;**18**:5687–93.
52. Wang YC, Zhang CH, Deng NY, *et al*. Kernel-based data fusion improves the drug-protein interaction prediction. *Comput Biol Chem* 2011;**35**:353–62.
53. 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.
54. Cao DS, Liu S, Xu QS, *et al*. Large-scale prediction of drug-target interactions using protein sequences and drug topological structures. *Anal Chim Acta* 2012;**752**:1–10.
55. Gonen M. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics* 2012;**28**:2304–10.
56. Tabei Y, Pauwels E, Stoven V, *et al*. Identification of chemogenomic features from drug-target interaction networks using interpretable classifiers. *Bioinformatics* 2012;**28**:i487–94.
57. Tabei Y, Yamanishi Y. Scalable prediction of compound-protein interactions using minwise hashing. *BMC Syst Biol* 2013;**7**(Suppl 6):S3.
58. Nanni L, Lumini A, Brahnam S. A set of descriptors for identifying the protein-drug interaction in cellular networking. *J Theor Biol* 2014;**359**:120–8.
59. Yang F, Xu J, Zeng J. Drug-target interaction prediction by integrating chemical, genomic, functional and pharmacological data. *Pac Symp Biocomput* 2014;148–59.
60. Mousavian Z, Masoudi-Nejad A. Drug-target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metab Toxicol* 2014;**10**:1273–87.
61. Liu H, Sun J, Guan J, *et al*. Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics* 2015;**31**:i221–9.
62. Pahikkala T, Airola A, Pietila S, *et al*. Toward more realistic drug-target interaction predictions. *Brief Bioinform* 2015;**16**:325–37.
63. Shi JY, Yiu SM, Li YM, *et al*. Predicting drug-target interaction for new drugs using enhanced similarity measures and super-target clustering. *Methods* 2015;**83**:98–104.
64. Ezzat A, Wu M, Li XL, *et al*. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinform* 2016;**17**:509.
65. Fu G, Ding Y, Seal A, *et al*. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinform* 2016;**17**:160.
66. Hao M, Wang Y, Bryant SH. Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique. *Anal Chim Acta* 2016;**909**:41–50.
67. Li ZC, Huang MH, Zhong WQ, *et al*. Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics* 2016;**32**:1057–64.
68. Mousavian Z, Khakabimamaghani S, Kavousi K, *et al*. Drug-target interaction prediction from PSSM based evolutionary information. *J Pharmacol Toxicol Methods* 2016;**78**:42–51.
69. Ozturk H, Ozkirimli E, Ozgur A. A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. *BMC Bioinform* 2016;**17**:128.
70. 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**:i18–27.
71. Ding YJ, Tang JJ, Guo F. Identification of drug-target interactions via multiple information integration. *Inf Sci* 2017;**418**:546–60.
72. Ezzat A, Wu M, Li XL, *et al*. Drug-target interaction prediction using ensemble learning and dimensionality reduction. *Methods* 2017;**129**:81–8.
73. Jiang J, Wang N, Chen P, *et al*. DrugECs: an ensemble system with feature subspaces for accurate drug-target interaction prediction. *Biomed Res Int* 2017;**2017**:6340316.
74. Keum J, Nam H. SELF-BLM: prediction of drug-target interactions via self-training SVM. *PLoS One* 2017;**12**:e0171839.
75. Li Z, Han P, You ZH, *et al*. In silico prediction of drug-target interaction networks based on drug chemical structure and protein sequences. *Sci Rep* 2017;**7**:11174.
76. Meng FR, You ZH, Chen X, *et al*. Prediction of drug-target interaction networks from the integration of protein sequences and drug chemical structures. *Molecules* 2017;**22**:1119.
77. 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**:17731.
78. Zhang J, Zhu MC, Chen P, *et al*. DrugRPE: random projection ensemble approach to drug-target interaction prediction. *Neurocomputing* 2017;**228**:256–62.

79. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics* 2018;**34**:1164–73.
80. Sharma A, Rani R. BE-DTI': ensemble framework for drug target interaction prediction using dimensionality reduction and active learning. *Comput Methods Prog Biomed* 2018;**165**:151–62.
81. 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*. 2018;**111**: 1839–1852.
82. Wang L, You ZH, Chen X, *et al*. RFDT: a rotation Forest-based predictor for predicting drug-target interactions using drug structure and protein sequence information. *Curr Protein Pept Sci* 2018;**19**:445–54.
83. Li Y, Huang YA, You ZH, *et al*. Drug-target interaction prediction based on drug fingerprint information and protein sequence. *Molecules* 2019;**24**:2999.
84. Mahmud SMH, Chen WY, Jahan H, *et al*. iDTi-CSsmoteB: identification of drug-target interaction based on drug chemical structure and protein sequence using XGBoost with over-sampling technique SMOTE. *IEEE Access* 2019;**7**:48699–714.
85. Rayhan F, Ahmed S, Md Farid D, *et al*. CFSBoost: cumulative feature subspace boosting for drug-target interaction prediction. *J Theor Biol* 2019;**464**:1–8.
86. Xuan P, Sun C, Zhang T, *et al*. Gradient boosting decision tree-based method for predicting interactions between target genes and drugs. *Front Genet* 2019;**10**:459.
87. Chen H, Zhang Z. A semi-supervised method for drug-target interaction prediction with consistency in networks. *PLoS One* 2013;**8**:e62975.
88. Gu Q, Ding YS, Zhang TL, *et al*. Prediction drug-target interaction networks based on semi-supervised learning method. In: *Proceedings of the 35th Chinese Control Conference 2016*, 2016, 7185–8.
89. Zhu S, Okuno Y, Tsujimoto G, *et al*. A probabilistic model for mining implicit 'chemical compound-gene' relations from literature. *Bioinformatics* 2005;**21**(Suppl 2):ii245–51.
90. Wang Y, Zeng J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics* 2013;**29**:i126–34.
91. Xie L, He S, Song X, *et al*. Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genomics* 2018;**19**:667.
92. Wen M, Zhang Z, Niu S, *et al*. Deep-learning-based drug-target interaction prediction. *J Proteome Res* 2017;**16**: 1401–9.
93. Tian K, Shao M, Wang Y, *et al*. Boosting compound-protein interaction prediction by deep learning. *Methods* 2016;**110**:64–72.
94. Lee I, Keum J, Nam H. DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput Biol* 2019;**15**:e1007129.
95. 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**:104–11.
96. Su R, Liu X, Wei L, *et al*. Deep-resp-forest: a deep forest model to predict anti-cancer drug response. *Methods* 2019;**166**:91–102.
97. Guo Y, Liu S, Li Z, *et al*. BCDForest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. *BMC Bioinform* 2018;**19**: 118.
98. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.
99. Schomburg I, Chang A, Ebeling C, *et al*. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 2004;**32**:D431–3.
100. Gunther S, Kuhn M, Dunkel M, *et al*. SuperTarget and matador: resources for exploring drug-target relationships. *Nucleic Acids Res* 2008;**36**:D919–22.
101. Wishart DS, Knox C, Guo AC, *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;**34**:D668–72.
102. Leslie C, Eskin E, Noble WS. The spectrum kernel: a string kernel for SVM protein classification. *Pac Symp Biocomput* 2002;564–75.
103. Kuhn M, Letunic I, Jensen LJ, *et al*. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016;**44**:D1075–9.
104. Takarabe M, Kotera M, Nishimura Y, *et al*. Drug target prediction using adverse event report systems: a pharmacogenomic approach. *Bioinformatics* 2012;**28**:i611–8.
105. 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.
106. Zhou ZH, Feng J. Deep forest: Towards an alternative to deep neural networks, arXiv preprint 2017; arXiv: 1702.08835.
107. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
108. Ding J, Li X, Hu H. TarPmiR: a new approach for microRNA target site prediction. *Bioinformatics* 2016;**32**:2768–75.
109. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA: ACM, 2016, 785–94.
110. Wadhwa S, Gupta A, Dokania S, *et al*. A hierarchical anatomical classification schema for prediction of phenotypic side effects. *PLoS One* 2018;**13**:e0193959.
111. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform* 2017;**18**:851–69.
112. Qu J, Chen X, Sun YZ, *et al*. Inferring potential small molecule-miRNA association based on triple layer heterogeneous network. *J Chem* 2018;**10**:30.
113. Chen X, Guan NN, Sun YZ, *et al*. MicroRNA-small molecule association identification: from experimental results to computational models. *Brief Bioinform* 2018.
114. Yin J, Chen X, Wang CC, *et al*. Prediction of small molecule-MicroRNA associations by sparse learning and heterogeneous graph inference. *Mol Pharm* 2019;**16**:3157–66.
115. Qu J, Chen X, Sun YZ, *et al*. In Silico prediction of small molecule-miRNA associations based on the HeteSim algorithm. *Mol Ther Nucleic Acid* 2019;**14**:274–86.
116. Wang CC, Chen X, Qu J, *et al*. RFSMMA: a new computational model to identify and prioritize potential small molecule-MiRNA associations. *J Chem Inf Model* 2019;**59**:1668–79.
117. Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 2018;**34**:i821–9.
118. Karimi M, Wu D, Wang Z, *et al*. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* 2019;**35**:3329–38.