


METHODOLOGY ARTICLE

Open Access



# Drug-target interaction prediction with tree-ensemble learning and output space reconstruction

Konstantinos Pliakos<sup>1,2\*</sup>  and Celine Vens<sup>1,2</sup>

## Abstract

**Background:** Computational prediction of drug-target interactions (DTI) is vital for drug discovery. The experimental identification of interactions between drugs and target proteins is very onerous. Modern technologies have mitigated the problem, leveraging the development of new drugs. However, drug development remains extremely expensive and time consuming. Therefore, *in silico* DTI predictions based on machine learning can alleviate the burdensome task of drug development. Many machine learning approaches have been proposed over the years for DTI prediction. Nevertheless, prediction accuracy and efficiency are persisting problems that still need to be tackled. Here, we propose a new learning method which addresses DTI prediction as a multi-output prediction task by learning ensembles of multi-output bi-clustering trees (eBICT) on reconstructed networks. In our setting, the nodes of a DTI network (drugs and proteins) are represented by features (background information). The interactions between the nodes of a DTI network are modeled as an interaction matrix and compose the output space in our problem. The proposed approach integrates background information from both drug and target protein spaces into the same global network framework.

**Results:** We performed an empirical evaluation, comparing the proposed approach to state of the art DTI prediction methods and demonstrated the effectiveness of the proposed approach in different prediction settings. For evaluation purposes, we used several benchmark datasets that represent drug-protein networks. We show that output space reconstruction can boost the predictive performance of tree-ensemble learning methods, yielding more accurate DTI predictions.

**Conclusions:** We proposed a new DTI prediction method where bi-clustering trees are built on reconstructed networks. Building tree-ensemble learning models with output space reconstruction leads to superior prediction results, while preserving the advantages of tree-ensembles, such as scalability, interpretability and inductive setting.

**Keywords:** Drug-target networks, Network reconstruction, Interaction prediction, Tree-ensembles, multi-output prediction

## Background

Predicting accurately drug-target interactions (DTI) is vital for the development of new drugs. Accurate and efficient identification of interactions between drugs and target proteins can accelerate the drug development process and reduce the required cost. In addition, the identification of drug-target interactions can unveil hidden drug or protein functions and shed light to enigmatic disease pathology mechanisms [1]. It can also provide scientists with insights which help in foreseeing adverse effects of

drugs [2, 3]. Furthermore, apart from discovering new drugs, DTI prediction can also leverage drug repositioning [2, 4–6], which aims at revealing new uses for already approved drugs. However, despite the persisting efforts made by the scientific community, experimentally identifying DTIs remains extremely demanding in terms of both time and expenses [7, 8]. The employment of computational methods and especially machine learning for *in silico* DTI prediction is thereby crucial for drug discovery and repositioning. Machine learning models can direct experiments, reveal latent patterns in large scale drug or protein data collections and extract unprecedented knowledge in drug-target networks.

\*Correspondence: [konstantinos.pliakos@kuleuven.be](mailto:konstantinos.pliakos@kuleuven.be)

<sup>1</sup>KU Leuven, Campus KULAK, Faculty of Medicine, Kortrijk, Belgium

<sup>2</sup>ITEC, imec research group at KU Leuven, Kortrijk, Belgium



Machine learning has shown great potential when employed in medicine and bioinformatics, especially in prediction or clustering tasks [9–11]. The most appealing field of machine learning is the supervised learning, where the learning models are constructed on an input set  $\mathcal{X}$  and an output set  $\mathcal{Y}$ , ( $f : \mathcal{X} \rightarrow \mathcal{Y}$ ). The instances (e.g., drugs, proteins) are represented by a set of feature vectors and they are also associated with an output variable. The goal is the learning of a function, based on the features of a training set of instances, which predicts the output [12]. In inductive modelling, when this function (model) is built, one can employ it to predict the output of new instances. The task is called regression in cases where the output is numeric and classification when it is categorical.

### Multi-output prediction in drug discovery

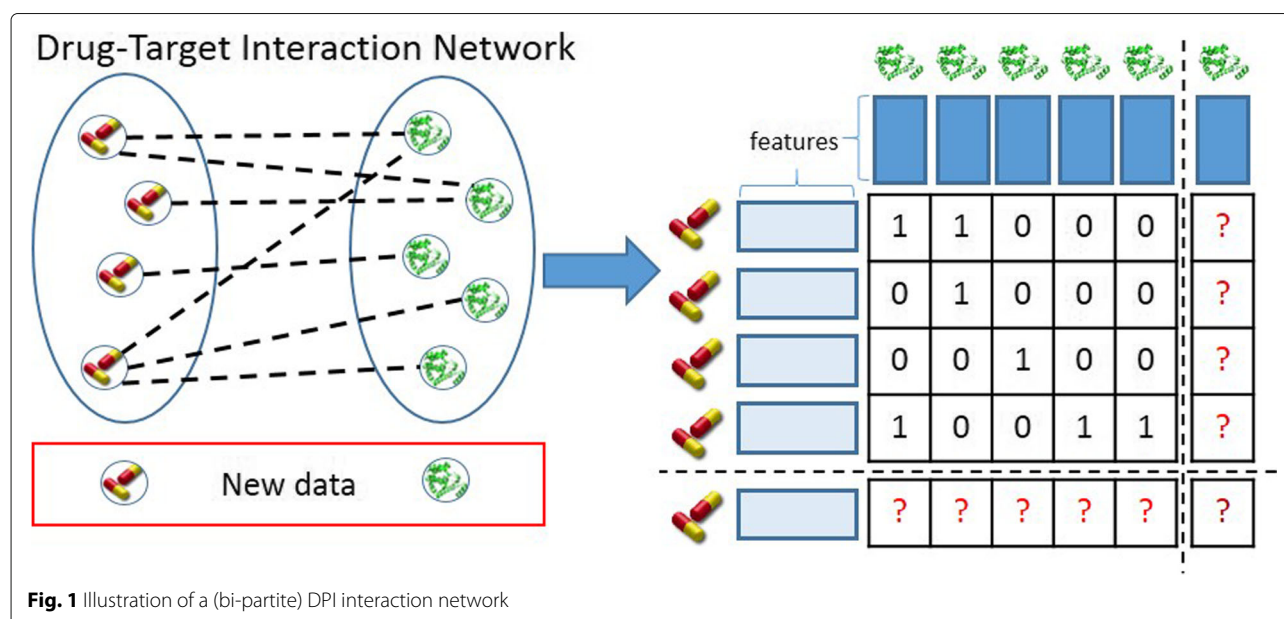
An interesting extension of typical classification or regression problems is the task of multi-output (multi-target) prediction [13]. In this case, the model learns to predict multiple output variables at the same time. Subcategories of multi-target prediction are multi-target classification (i.e., the targets have categorical values) and multi-target regression [14]. A distinctive condition is multi-label classification [15, 16]. This can be translated as multi-target regression with only zero and one as numeric values for each target, or as multi-target classification, with only binary values for each target.

Multi-output prediction models learn from multiple outputs simultaneously. They are often benefited from exploiting possible correlations between the targets, improving this way their prediction performance. In particular, when it comes to drug discovery, the interest in

multi-output models is even greater. In the past, the learning methods proposed for DTI prediction aimed at performing predictions for a specific target protein, admitting the old paradigm of ‘one target, one drug, one disease’. This strategy led to inferior performance as the drug-disease relation complexity is far greater [17, 18]. The majority of known diseases are usually associated with multiple proteins [19]. It has been generally admitted that drugs which interact with multiple target proteins (*polypharmacology*) are more effective [20–22]. Multi-output learning can also contribute to investigating the off-target drug activity (i.e., unintended function of a drug). The investigation of such activities can lead to new uses for existing drugs (drug repositioning) or contrarily, the identification of unwanted side-effects. Such adverse reactions of drug candidates are usually identified at a later stage of the drug development process, leading to extremely expensive late stage failures.

### DTI networks

A drug-protein interaction network is a heterogeneous network (also referred to as bi-partite graph) that can be formulated as a collection of two sets of items that interact with each other. Each item set is described by its own features which compose the background information in our problem. The interactions are the links connecting the nodes of the network and are often represented as a matrix, often denoted as interaction, adjacency, or connectivity matrix. In this paper, we use the term interaction matrix. In Fig. 1, an illustration of a DTI network in the aforementioned setting is displayed. One can follow two learning strategies in this framework: the local [23] and



the global [24]. A discussion of these two strategies took place originally in [25] and later in [26, 27].

Traditional DTI prediction models based on the local approach handle the two sets of the DTI network separately. In particular, they first divide the DTI network into different (traditional) feature sets, the drug-based set and the protein-based one. Next, each set's learning task is tackled separately and then the results are combined. Often, in the absence of information on both sides, local models are built on a single feature space, ligand (drug) space or target protein space. Ligand-based models are built on the known ligands that interact with the target proteins. However, the performance of these models is impaired when it comes to target proteins with only a really small number (or even none) of known binding ligands [28]. Alternatively, target-based models are built on the target proteins using protein (3D) structure information. Nevertheless, the 3D structure of many target proteins is often unavailable.

Due to these bottlenecks, the interest of the scientific community was shifted towards a global setting referred to as *chemogenomics* [29, 30]. The underlying idea behind the global setting is that drug information is integrated with target protein information and thereby complement each other. However, this setting also suffers from weaknesses. Global approaches are mostly based on matrix factorization or graph learning, following the transductive setup (i.e., the test instances are needed in the training phase). Alternatively, there are other approaches which are based on inductive classifiers. In these cases, DTI prediction is treated as a binary classification problem where classifiers are trained over the Cartesian product of drug-related and target-related feature sets. This Cartesian product often leads to an enormous data matrix. Thus, these approaches are computationally very expensive and not particularly scalable. Furthermore, in this global setting, one assumes that rich background information (feature vectors) is always available for both all drugs and all their targets, which is not always the case. Despite these disadvantages, global approaches remain the most promising.

### Introduction to the proposed method

Major problems in DTI prediction are the present noise in the output space, the existence of no true negative interactions and the extreme class imbalance. These problems are not easily surpassed and they often devastate the predictive performance of even powerful learning methods. There is a plethora of studies aiming at feature space transformation, removing noise or revealing latent manifolds in the data. However, to the best of our knowledge, there is almost nothing on integrating supervised learning methods with output space reconstruction. An intelligent

reconstruction can remove the existing noise, reveal latent patterns and mitigate class imbalance in the output space.

In this paper, we propose a new DTI prediction framework that provides great predictive performance while being computationally efficient and scalable. We propose that building multi-output learning models on reconstructed networks leads to superior predictive performance. Our approach addresses DTI prediction as a multi-output prediction task, building tree-ensemble learning models and specifically ensembles of bi-clustering trees (eBICT) [27, 31], on reconstructed networks. Although other inductive learning models could have been employed, we designate eBICT because it inherits the merits of tree-ensembles, such as scalability, computational efficiency, and interpretability. eBICT also provides bi-clustering [32] of the interaction matrix as a side product.

Reconstructing a DTI network is a challenging problem and various approaches have been proposed over the years. The most effective approaches are typically related to matrix factorization. Scientists have extended the traditional optimization problem of matrix factorization including multiple constraints. Recently, a neighborhood regularized logistic matrix factorization (NRLMF) [33] method was presented, integrating logistic matrix factorization (LMF) with neighborhood regularization taking also into account class imbalance. The authors obtained outstanding results, naming their method a state of the art in DTI prediction. Here, we employ NRLMF for reconstructing the target space in our problem and we show that the predictive performance of inductive learning models is particularly boosted when they are integrated with output space reconstruction. The proposed multi-output prediction framework combines great prediction performance with scalability, computational efficiency, and interpretability. The proposed method offers bi-clustering of a drug-target network as a side product and also follows the inductive setup. The latter means that neither the test instances are needed in the training process nor the training instances are required to perform predictions for new instances. Furthermore, the proposed method is apt to perform predictions for new candidate drugs, a setting more applied to drug discovery, new target proteins, a setting more applied to drug repositioning, or new drug-protein pairs.

### Related work

Recently, great interest has been witnessed in developing machine learning models for DTI prediction [34]. Kernel learning was employed for DTI prediction in [35], where the authors constructed kernels for drugs, target proteins and the interaction matrix. DTI prediction was then performed using the regularized least squares classifier. This approach was later extended to handle new candidate

drugs or target proteins in [36]. In [37], a semi-supervised approach was proposed integrating similarities between drugs and local correlations between targets into a robust PCA model. Deep learning strategies for DTI prediction were used in [38, 39]. An interesting multi-label classification framework exploiting label partitioning was recently proposed for DTI prediction in [40] as well as in the 7th chapter of [41]. Furthermore, the authors in [42] employed multi-domain manifold learning and semidefinite programming for DTI prediction while in [43] it was handled using label propagation with linear neighborhood information. Moreover, Shi et al. [44] presented an MLkNN [45] driven approach to predict interactions between new candidate drugs and target proteins. The method was based on clustering the features of the target proteins. A second interaction matrix was constructed based on this super-target clustering. The MLkNN was applied to both interaction matrices and final predictions were yielded as an integration of the individual prediction scores. MLkNN was also used in [46] for drug side effect prediction. A feature selection-based MLkNN method was presented, which combined the construction of multi-label prediction models with the determination of optimal dimensions for drug-related feature vectors.

Many promising predictors were based on matrix factorization [30]. For instance, in [47], graph regularization was incorporated into matrix factorization. In particular, the proposed method consisted of two steps. First, a weighted k Nearest Neighbor (k-NN) was employed, converting the binary interaction scores into numeric ones. Next, a graph regularization driven matrix factorization method was applied. In [33], the authors proposed a neighborhood regularized logistic matrix factorization (NRLMF) approach. Their method incorporated neighborhood regularization into logistic matrix factorization. The performance of their approach was also enhanced by applying a weighing scheme that favored the pairs where an interaction occurs. In [29], another similar extension to logistic matrix factorization (LMF) was presented. The authors integrated LMF with multiple kernel learning and graph Laplacian regularization.

Extensive work has been also noted in building ensemble learning models. In more detail, a synergistic model was built in [28]. It achieved a fair predictive performance integrating predictions from multiple methods into a Learning to Rank framework. In [48], ensemble learning was also used along with strategies tackling existing class-imbalance in drug-target networks.

Moreover, several approaches emphasized on transforming or extending the feature space, generating more informative representations of the DTI network. Next, the final predictions were yielded as the output of a common classifier. In [49], the authors used network (graph) mining to extract features. Next, a Random Forest (RF)

[50] classifier was applied to predict the interactions. Similarly in [51], the authors exploited the topology of the DTI network to extract features. The final predictions were performed using a Random Forest classifier. In addition, Liu et al. [52] proposed a strategy to identify highly negative samples before applying a classifier.

## Results

### Evaluation metrics

In order to evaluate the proposed approach we employed two metrics in a micro-average setup, namely area under the receiver operating characteristic curve (AUROC) and area under precision recall curve (AUPR). ROC curves correspond to the true positive rate ( $\frac{TP}{TP+FN}$ ) against the false positive rate ( $\frac{FP}{FP+TN}$ ) at various thresholds. Precision-Recall curves correspond to the Precision ( $\frac{TP}{TP+FP}$ ) against the Recall ( $\frac{TP}{TP+FN}$ ) at various thresholds.

In Table 3 it can be seen that the interaction datasets are very sparse, which makes the corresponding classification task very class imbalanced. Generally, AUPR is considered more informative than AUROC in highly imbalanced classification problems [53, 54]. Nevertheless, it is important to note that in drug discovery the crucial value is to minimize the false negatives (FN), these are interactions which are positive but overlooked by the computational predictor. Any positive in silico predictions will get validated in the lab, whereas strong negative ones are rarely checked.

### Evaluation protocol

A major point in our paper is to evaluate the contribution of output space reconstruction to the predictive performance of multi-output learning models. To this end, our evaluation study begins with comparing the proposed DTI approach (BICTR) against ensemble of bi-clustering trees (eBICT) without output space reconstruction. Next, we compare BICTR to three state of the art DTI prediction methods, BLMNII [36], STC [44], and NRLMF [33]. The method in [36] is denoted as BLMNII and is a kernel-based local approach. The method in [44] is denoted as super target clustering (STC). It uses MLkNN in a target clustering-driven strategy. The methods are compared in the three prediction settings presented in the “Method” section, namely  $T_d \times L_p$ ,  $L_d \times T_p$ , and  $T_d \times T_p$ . We performed comparisons independently for every setting. Both BLMNII and STC are local models and the predictions between pairs of new drugs and new targets were performed following the standard two step approach proposed in [26, 55].

In  $T_d \times L_p$  and  $L_d \times T_p$  we used 10-fold cross validation (CV) on nodes (i.e., CV on drugs and CV on targets, respectively). It is important to clarify that when a drug  $d_i$  is included in the test set of the  $T_d \times L_p$  setting the whole interaction profile of  $d_i$  should not be present in



the training set. The same holds for the target proteins in the  $L_d \times T_p$  setting. In  $T_d \times T_p$ , we used CV on blocks of drugs and targets. For every iteration, we removed one fold corresponding to drugs and one fold corresponding to proteins from the learning set and used their combined interactions as test set. When a drug-target pair  $(d_i, p_j)$  is included in the test set this means that the whole interaction profile of both  $d_i$  and  $p_j$  should not be present in the training set. In  $T_d \times T_p$ , we used 5-fold CV over blocks of drugs and targets (i.e.,  $5 \times 5 = 25$  folds). This was done because the data are very sparse and the application of a 10-fold CV setting was difficult.

The number of trees in tree-ensemble algorithms was set to 100 without tree-pruning. The parameter  $c$  in Eq. 2, which defines the weight of the positive (interacting) drug-target pairs, was set equal to 5 as in [33]. All the other parameters of NRLMF, shown in Eq. 2, were optimized in a 5-fold CV inner tuning process (nested CV) following a grid search. More specifically, parameters  $\lambda_d, \lambda_p, \alpha, \beta$  as well as the optimal learning rate were selected from a range of  $\{2^{-2}, 2^{-1}, 2^0, 2^1\}$ . The number of nearest neighbors was selected from  $\{3, 5, 10\}$  and the number of latent factors from  $\{50, 100\}$ . For BLMNII, we used the rbf kernel as proposed in the corresponding paper and tuned the linear combination weight through 5-fold CV inner tuning (nested CV), picking values in  $\{0.1, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5\}$ . The number of nearest neighbors in STC was also tuned through 5-fold CV inner tuning (nested CV), picking values in  $\{3, 5, 7, 9, 11\}$ .

### Obtained results

The AUROC and AUPR results are presented in Tables 1 and 2, respectively. Best results are shown in bold faces and \* indicates that the results between BICTR and its competitor were found statistically significantly different ( $p < 0.05$ ) based on a Wilcoxon Signed-Ranks Test run on the CV-folds. As it is reflected, BICTR outperforms eBICT in all three prediction settings, in terms of both AUROC and AUPR. Specifically, BICTR significantly outperforms eBICT in every dataset in terms of AUROC. It also achieves better AUPR results in every dataset and setting. The only exceptions occur in the E dataset in  $T_d \times L_p$  and  $T_d \times T_p$  where nonetheless the differences are not statistically significant. Thus, the original hypothesis that network reconstruction can boost the predictive performance of multi-output learning models is verified.

We next evaluated BICTR by comparing it against state of the art DTI prediction approaches and the obtained AUROC and AUPR results are also presented in Tables 1 and 2, respectively. BICTR overall outperforms its competitors, affirming its effectiveness in DTI prediction. More specifically, BICTR surpasses BLMNII and STC in all prediction settings, both in terms of AUROC and AUPR. When it comes to NRLMF, BICTR yields better

**Table 1** AUROC results for the compared methods

AUROC					
Data	BICTR	eBICT	NRLMF	BLMNII	STC
$T_d \times L_p$					
NR	<b>0.875</b>	0.787*	0.851*	0.807*	0.794*
GR	<b>0.894</b>	0.857*	0.867*	0.842*	0.847*
IC	<b>0.811</b>	0.780*	0.792	0.737*	0.783*
E	<b>0.891</b>	0.827*	0.777*	0.815*	0.794*
Avg	<b>0.868</b>	0.813	0.822	0.800	0.805
$L_d \times T_p$					
NR	<b>0.905</b>	0.614*	0.747*	0.667*	0.525*
GR	<b>0.951</b>	0.846*	0.861*	0.776*	0.800*
IC	<b>0.968</b>	0.931*	0.949*	0.887*	0.909*
E	<b>0.973</b>	0.924*	0.940*	0.904*	0.906*
Avg	<b>0.949</b>	0.829	0.874	0.809	0.785
$T_d \times T_p$					
NR	0.676	0.634*	0.683	0.554*	0.469*
GR	<b>0.811</b>	0.792*	0.800*	0.475*	0.630*
IC	<b>0.733</b>	0.719*	0.731	0.466*	0.649*
E	<b>0.812</b>	0.785*	0.749*	0.490*	0.682*
Avg	<b>0.758</b>	0.733	0.741	0.496	0.608

results in terms of AUROC in all settings and AUPR in  $L_d \times T_p$ . The AUPR results obtained by BICTR are inferior in  $T_d \times L_p$  and  $T_d \times T_p$ . Nevertheless, the differences are statistically significant only for the E dataset. In a case like that we could deduct that BICTR is better at maximizing true negatives (TN) while NRLMF is better at minimizing false positives (FP). In drug discovery the elimination of false positives, albeit important, is not as crucial as in other tasks because the possible hits or leads (i.e., positive interactions) will anyway get validated in the lab by (medicinal) chemists.

### Discussion

The obtained results indicate that output space reconstruction can elevate the performance of multi-output learning models, leading to more accurate DTI predictions. The effectiveness of BICTR was affirmed in all three DTI prediction settings. The contribution of the NRLMF-based step is substantial as it reconstructs the output space identifying potential non-reported drug-target interactions in the training set. This especially mitigates the problem of class imbalance. The performance improvement achieved by the output space reconstruction step was confirmed by conducted experiments, where BICTR clearly outperformed eBICT.

One could identify a connection between the approach presented in this chapter and the setting of Positive Unlabeled data (PU) learning [56]. Here, similar to PU learning, we acknowledge the lack of truly negative

**Table 2** AUPR results for the compared methods

AUPR					
Data	<i>BICTR</i>	<i>eBICT</i>	<i>NRLMF</i>	<i>BLMNII</i>	<i>STC</i>
$T_d \times L_p$					
NR	0.480	0.444*	<b>0.501</b>	0.436	0.467
GR	0.334	0.329	<b>0.349</b>	0.312	0.324
IC	<b>0.329</b>	0.317*	<b>0.329</b>	0.213*	0.307
E	0.314	0.316	<b>0.367*</b>	0.255*	0.353
Avg	0.364	0.352	<b>0.387</b>	0.304	0.363
$L_d \times T_p$					
NR	<b>0.555</b>	0.424*	0.485*	0.338*	0.384*
GR	<b>0.523</b>	0.504*	0.406*	0.324*	0.365*
IC	<b>0.808</b>	0.791*	0.798	0.724*	0.779*
E	0.785	0.784	<b>0.795*</b>	0.735*	0.752*
Avg	<b>0.668</b>	0.626	0.621	0.530	0.570
$T_d \times T_p$					
NR	0.160	0.151	<b>0.168</b>	0.100	0.080*
GR	0.156	0.156	<b>0.168</b>	0.042*	0.079*
IC	<b>0.233</b>	0.229	0.227	0.041*	0.198*
E	0.214	0.218	<b>0.263*</b>	0.016*	0.190*
Avg	0.191	0.189	<b>0.207</b>	0.050	0.137

drug-target pairs. In the first step of our approach (matrix factorization-based) we reconstruct the interaction matrix of the networks, identifying the likely positive (interacting) drug-target pairs from the set of unlabeled ones (zeros in the interaction matrix). The subsequent supervised learning method is applied on a reconstructed interaction matrix, which consists of zeros (i.e., strong negative drug-target pairs), ones (i.e., interacting drug-target pairs), and fuzzy values (i.e., ambiguous drug-target pairs).

It should be also highlighted that the proposed method follows the inductive setup as the reconstruction of the output space takes place only in the training process. This means that after the training process is complete, one can perform predictions for new data (e.g., new candidate drugs). In addition, the employed matrix factorization step does not affect the interpretability of tree-ensemble learning which is subsequently introduced in the proposed DTI prediction method.

Furthermore, different from other approaches (e.g., NRLME, STC, BLMNII), the proposed method does not require the training instances (feature vectors) to be kept, which can be vital for studies performed in large scale DTI networks. BICTR is not a similarity-based method and is perfectly applicable on other types of feature spaces. For example, one could use GO annotations or PFAM domains as protein related features and drug side effects or chemical compound interactions as drug-related features. Moreover, one could extract features from the network topology. In addition, as BICTR is a tree-ensemble

method, it adopts all the advantages of decision tree based learning. It is scalable, computationally efficient, interpretable, and capable of handling missing values.

Moreover, synergistic learning approaches that employ multiple classifiers to yield predictions are not considered as competitors. BICTR can be clearly integrated into such mechanisms. The performance of BICTR can be also boosted by feature construction methods based on graph embeddings. Finally, we state that although matrix factorization (NRLMF) was employed for reconstructing the output space, other approaches could be used as well.

## Conclusion

In this paper we have presented a new drug-target interaction prediction approach based on multi-output prediction with output space reconstruction. We showed that multi-output learning models can manifest superior predictive performance when built on reconstructed networks. Tree-ensemble learning models and specifically ensembles of bi-clustering trees were deployed in this framework, constructing an accurate and efficient DTI prediction method. The proposed approach was compared against state of the art DTI prediction methods on several benchmark datasets. The obtained results affirmed the merits of the proposed framework.

The learning method that was deployed here could be used to perform in silico predictions on large scale drug-target networks in the future. These predictions should get verified later in the lab, potentially revealing novel interactions.

## Method

In this section, we first discuss about the general structure of drug-target networks, present notations and describe different prediction settings. We then provide a broad description of tree-ensemble learning and multi-output prediction. Next, we present the individual mechanisms of bi-clustering trees and matrix factorization. Finally, the proposed DTI prediction approach is presented.

### Predicting drug-target interactions

Drug target interaction networks are heterogeneous networks, which are denoted as bi-partite graphs in graph theory. A DTI network consists of two finite sets of nodes  $D = \{d_1, \dots, d_{|D|}\}$  and  $P = \{p_1, \dots, p_{|P|}\}$ , that correspond to drugs and target proteins, respectively. Each node is represented by a feature vector. Drug-related features may consist of chemical structure similarities, drug side effects, or drug-drug interactions. Protein-related features may consist of protein sequence similarities, GO annotations, protein-protein interactions or protein functions. A link between two nodes of a DTI network corresponds to an existing interaction between the corresponding drug and target protein. The set of existing or not existing

network links form an interaction matrix  $\mathbf{Y} \in \mathbb{R}^{|D| \times |P|}$ . Every item  $y(i, j) \in \mathbf{Y}$  is equal to 1 if an interaction between items  $d_i$  and  $p_j$  exists and 0 otherwise.

DTI prediction, a task also denoted as DTI network inference, can be handled as a supervised learning task and especially as a classification task on pairs of nodes. The goal is to build a model that receives a drug-target pair as input and outputs a probability that an interaction between these two pair nodes holds. In the most practical inductive setup, the learning model is built on a training set of drug-target pairs and after the learning process is complete, it can perform predictions for unseen pairs.

One can perform DTI predictions for new drugs, new target proteins, or new drug-target pairs. The latter is clearly more challenging. Predicting interactions between drugs and targets that are both included in the training set is considered a semi-supervised learning task and is not studied in this paper as we focus on supervised learning. The addressed prediction framework is demonstrated in Fig. 2. The  $(L_d \times L_p)$  is the interaction matrix  $\mathbf{Y}$ . DTI prediction tasks can be divided in 3 settings.

- Test drugs - Learned targets ( $T_d \times L_p$ ): interactions between new drug candidates and target proteins that have been included in the learning procedure.
- Learned drugs - Test targets ( $L_d \times T_p$ ): interactions between drugs that have been included in the learning procedure and new target proteins.
- Test drugs - Test targets ( $T_d \times T_p$ ): interactions between new drug candidates and new target proteins.

The aforementioned prediction setting was thoroughly described in [26, 55, 57–59].

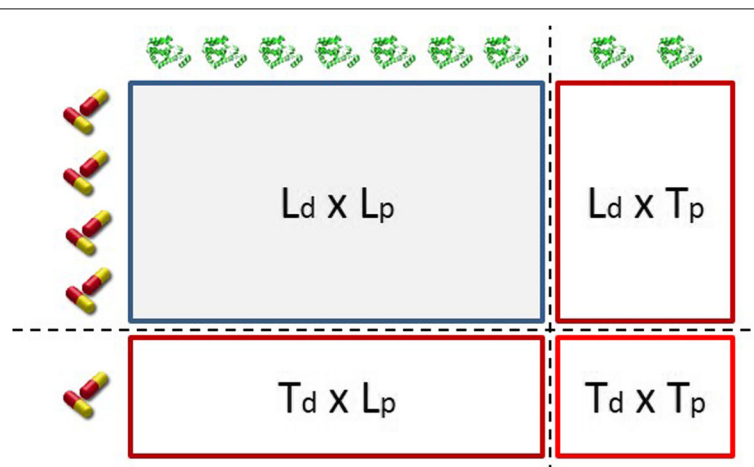
### Multi-output tree-ensembles

Decision tree induction algorithms [60] adopt a top-down architecture. The first node is called the root node. Every

node is recursively split after applying a test to one of the instance features. A split quality criterion (e.g., entropy, variance reduction etc.) is employed to measure the quality of the split. The best split is selected and the tree growing process continues until the data contained in a node is pure w.r.t. the labels. The tree growing can also stop if a stopping criterion is reached. The last nodes of the tree are called leaves. Every leaf receives a label, which is typically the average or the majority of the labels of the containing instances. A new (unseen) instance will traverse the tree and end up in a leaf node. The label that corresponds to this leaf is then given as a prediction to the new instance.

Single trees often suffer from the overfitting effect and are considered as relatively unstable models. However, when they are extended to tree-ensembles [50], they often achieve state-of-the-art performance. The overfitting effect is also tackled by tree-ensembles. Several tree-ensemble approaches exist. Two of the most popular and effective ones are the random forests (RF) [50] and the extremely randomized trees (ERT) [61]. Typically, it is more challenging to interpret a tree-ensemble model than a single tree-based one. Nevertheless, there are strategies [62] that transform a tree-ensemble to a single tree, avoiding this way the loss of the interpretability advantage. Another advantage of tree-ensembles is their ability to rank the features, based on their contribution to the learning procedure. Although the predictive performance of tree-ensembles may slightly vary based on the different randomization seeds, they are considered as very stable predictors.

Moreover, most tree-based learning models can easily be applied to multi-output tasks, for example multi-label classification [63] or multi-target regression [14]. Multi-output models learn to predict multiple output variables simultaneously. In a DTI prediction task, the instances can



**Fig. 2** The prediction setting of a DTI network

be the drugs and the outputs (labels) are the drug-target interactions. When a new drug arrives, a set of labels is assigned to it. Each label of this set corresponds to an interaction between this drug and a target protein.

**Ensembles of bi-clustering trees**

Pliakos et al. [27] proposed a *bi-clustering tree* for interaction prediction, extending a single multi-output decision tree to the global network setting. That tree model is shown in Fig. 3 [27]. The model is built on pairs of instances and predicts the interactions between them. This method was then extended to the tree-ensemble setting in [31], utilizing the ERT mechanism. The trees grow having a random sub-set of both row and column features as split candidates, inducing therefore a bi-clustering of the network. A split on a row feature corresponds to a row-wise partitioning of the matrix while a split on a column-feature to a column-wise one. The final predictions are generated as the average of the predictions yielded by each one of the trees that form the ensemble collection.

**NRLMF**

In matrix factorization the goal is to compute two matrices that, when multiplied, approximate the input matrix. More concretely, in DTI prediction, the interaction matrix  $Y \in \mathbb{R}^{|D| \times |P|}$  is used as input and the task is to compute two matrices, namely  $U \in \mathbb{R}^{|D| \times k}$  and  $V \in \mathbb{R}^{|P| \times k}$ , so  $UV^T \approx Y$ . Matrices  $U$  and  $V$  are considered as  $k$ -dimensional latent representations of drugs and proteins, where  $k \ll |D|, |P|$ .

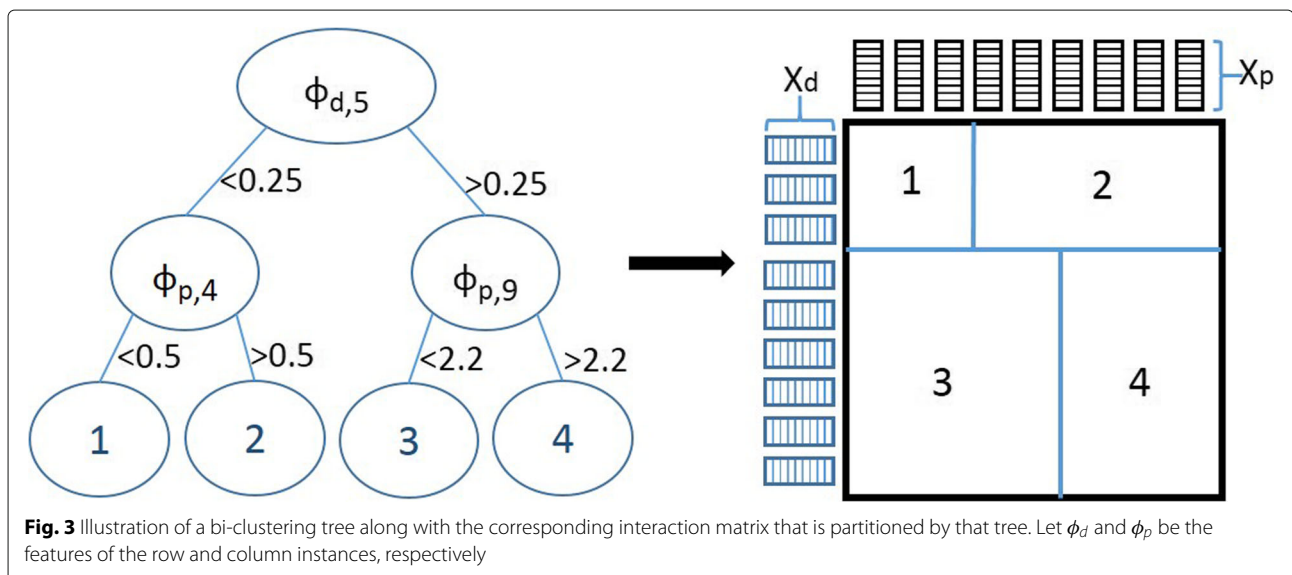
The Neighborhood Regularized Logistic Matrix Factorization (NRLMF) [33] is principally based on LMF, modelling the probability  $\hat{y}_{ij}$  that a drug  $d_i$  interacts with a target protein  $p_j$  as follows.

$$\hat{y}_{ij} = \frac{\exp(\mathbf{u}_i \mathbf{v}_j^T)}{1 + \exp(\mathbf{u}_i \mathbf{v}_j^T)} \tag{1}$$

The  $k$ -dimensional vectors  $\mathbf{u}_i$  and  $\mathbf{v}_j$  are latent representations of  $d_i$  and  $p_j$ , respectively. The original LMF expression is extended with two regularization terms which contribute to avoid overfitting and two graph regularization terms that capture the drug corresponding and protein corresponding neighborhood information. More thoroughly, the two regularization terms that appear in the second line of Eq. (2) stem from the application of zero-mean Gaussian priors on the latent vectors of all drugs and targets. They prevent overfitting by favoring simple solutions that consist of relatively small values. The next two terms are graph regularization terms that contribute to the optimization procedure by learning the underlying manifolds in the data. The final objective function that is yielded is shown below:

$$\begin{aligned} \min_{U, V} & \sum_{i=1}^{|D|} \sum_{j=1}^{|P|} (1 + cY_{ij} - Y_{ij}) \ln [1 + \exp(u_i v_j^T)] - cY_{ij} u_i v_j^T \\ & + \lambda_d \|U\|_F^2 + \lambda_p \|V\|_F^2 \\ & + \alpha \text{Tr}(U^T L^d U) + \beta \text{Tr}(V^T L^p V) \end{aligned} \tag{2}$$

Parameters  $\lambda_d$ ,  $\lambda_p$ ,  $\alpha$ , and  $\beta$  control the regularization terms while parameter  $c$  ( $c \geq 1$ ) expresses the weight of observed interacting drug-target pairs to the optimization process. The idea was that these interacting pairs have been experimentally verified and are therefore more important than unknown pairs (i.e.,  $Y_{ij} = 0$ ). By adjusting  $c$ , we specify the importance level of interacting pairs to the optimization process. Moreover, when  $c > 1$  each interaction pair is treated as  $c$  positive pairs. This contributes to the mitigation of the class-imbalance problem.



**Fig. 3** Illustration of a bi-clustering tree along with the corresponding interaction matrix that is partitioned by that tree. Let  $\phi_d$  and  $\phi_p$  be the features of the row and column instances, respectively



### Bi-clustering trees with output space reconstruction

In our DTI task we assume that there are originally no truly negative drug-target pairs but only positive and unlabeled ones, which can be either positive (not reported yet) or negative. This setting is often referred to as Positive-Unlabeled (PU) learning setting [56]. The proposed approach learns bi-clustering trees with output space reconstruction (BICTR). This way tree-ensemble learning, a powerful supervised learning family of algorithms, is integrated with semi-supervised driven approaches, such as matrix factorization. Here, we promote ensembles of bi-clustering trees and NRLMF.

We first reconstruct the output space, exploiting neighborhood information, revealing underlying manifolds in the topology of the DTI network (i.e., interaction matrix) and alleviating class-imbalance. The input of our approach is the drug-related feature space  $\mathbf{X}_d$ , the target-related feature space  $\mathbf{X}_p$ , and the interaction matrix  $\mathbf{Y}$ . We reconstruct the DTI network by learning matrices  $\mathbf{U}$  and  $\mathbf{V}$  based on Eq. 2. The new interaction matrix is denoted as  $\hat{\mathbf{Y}}$  and every  $\hat{y}_{ij} \in \hat{\mathbf{Y}}$  is computed as in Eq. 1. Although actually interacting pairs of the network have already received an increased level of importance through the reconstruction process, we support even further the verified interactions as follows:

$$\hat{y}_{ij} = \begin{cases} 1, & \text{if } y_{ij} = 1 \\ \hat{y}_{ij}, & \text{otherwise.} \end{cases} \quad (3)$$

Next, we learn eBICT on the reconstructed target space. In more detail, the input for every tree in our ensemble is the drug-related feature space  $\mathbf{X}_d$ , the target-related feature space  $\mathbf{X}_p$ , and the reconstructed interaction matrix  $\hat{\mathbf{Y}}$ . The root node of every tree in our setting contains the whole interaction network and a partitioning of this network is conducted in every node. The tree growing process is based on both vertical and horizontal splits of the reconstructed interaction matrix  $\hat{\mathbf{Y}}$ . The variance reduction is computed as  $Var = \sum_j^{|\mathcal{P}|} Var(\hat{\mathbf{Y}}_j)$  when the split test is on  $\phi_d \in \mathbf{X}_d$  and  $Var = \sum_i^{|\mathcal{D}|} Var(\hat{\mathbf{Y}}_i^T)$  when the split test is on a  $\phi_p \in \mathbf{X}_p$ .

The NRLMF-based target space reconstruction step of the proposed DTI prediction strategy boosts the predictive performance of the eBICT while preserving all the advantages of tree-ensembles, such as scalability, computational efficiency, and interpretability. An analysis of the computational efficiency and interpretability of bi-clustering trees took place in [27]. The approach that is proposed here, despite being integrated with matrix factorization, continues to follow the inductive setup. In more detail, the output space reconstruction process takes place only in the training process. After the training model is complete, new instances that may arrive (e.g., new candidate drugs) just traverse the grown bi-clustering trees

**Table 3** The drug-protein networks (DPN) used in the experimental evaluation are presented

DPN	drugs  ×  proteins	Features	interactions
NR	54 × 26	54 – 26	90/1404 (6.4%)
GR	223 × 95	223 – 95	635/21185 (3%)
IC	210 × 204	210 – 204	1476/42840 (3.4%)
E	445 × 664	445 – 664	2926/295480 (1%)

and predictions are assigned to them based on the leaves in which they end up.

### Data

We employed 4 benchmark datasets that represent drug-target interaction networks [64]. The characteristics of each network are shown in Table 3. More specifically, this table contains the number of drugs, proteins, and existing interactions in every network. The number of features used to represent each sample (drug or protein) is also displayed.

The datasets in [64] correspond to 4 drug-target interaction networks where the interactions between drugs and target proteins are represented as binary values. In these networks, compounds interact with proteins that belong to 4 pharmaceutically useful categories: nuclear receptors (NR), G-protein-coupled receptors (GR), ion channels (IC), and enzymes (E). The features that describe the drugs are similarities based on their chemical structure. The features representing the target proteins correspond to similarities based on the alignment of protein sequences. The sequence similarities were calculated according to the normalized Smith-Waterman score.

### Consent to publish

Not applicable

### Abbreviations

AUPR: Area under precision recall curve; AUROC: Area under the receiver operating characteristic curve; BICTR: Bi-Clustering Trees with output space Reconstruction; DPI: Drug-protein interaction; DTI: drug-target interaction; E: Enzymes; eBICT: Ensemble of bi-clustering trees; ERT: Extremely randomized trees; GR: G-protein-coupled receptors; IC: Ion channels; LMF: Logistic matrix factorization; MLKNN: Multi-label k-nearest neighbor; NR: Nuclear receptors; NRLMF: Neighborhood regularized logistic matrix factorization; RF: Random forests; STC: Super target clustering

### Acknowledgements

The authors would like to thank KU Leuven and imec for the financial support.

### Authors' contributions

KP developed the method and ran the experiments. CV provided supervision and valuable advice in all parts of this study. Both authors contributed to the text. All authors read and approved the final manuscript.

### Funding

The authors would like to thank KU Leuven and imec for the financial support. The funding did not play any role in the design of the study, collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The data and materials used in this study can be found here: <http://www.montefiore.ulg.ac.be/~schrynemackers/datasets>, <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/>, <https://github.com/gerdinard/myprojects/>

**Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 21 June 2019 Accepted: 21 January 2020

Published online: 07 February 2020

**References**

- Núñez S, Venhorst J, Kruse C. G. Target-drug interactions: first principles and their application to drug discovery. *Drug Discov Today*. 2012;17(1-2):10–22. <https://doi.org/10.1016/J.DRUDIS.2011.06.013>.
- Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kujjer MB, Matos RC, Tran TB, Whaley R, Glennon RA, Hert J, Thomas KLH, Edwards DD, Shoichet BK, Roth BL. Predicting new molecular targets for known drugs. *Nature*. 2009;462(7270):175–81. <https://doi.org/10.1038/nature08506>.
- Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, Lavan P, Weber E, Doak AK, Côté S, Shoichet BK, Urban L. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*. 2012;486(7403):361–7. <https://doi.org/10.1038/nature11159>.
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004;3(8):673–83. <https://doi.org/10.1038/nrd1468>.
- Wu Z, Cheng F, Li J, Li W, Liu G, Tang Y. SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief Bioinforma*. 2016;18(2):012. <https://doi.org/10.1093/bib/bbv012>.
- Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. *Brief Bioinforma*. 2016;17(1):2–12. <https://doi.org/10.1093/bib/bbv020>.
- Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203–14. <https://doi.org/10.1038/nrd3078>.
- Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: A systematic review. *Health Policy*. 2011;100(1):4–17. <https://doi.org/10.1016/j.healthpol.2010.12.002>.
- Tarca AL, Carey VJ, Chen X-w, Romero R, Drăghici S. Machine Learning and Its Applications to Biology. *PLoS Comput Biol*. 2007;3(6):116. <https://doi.org/10.1371/journal.pcbi.0030116>.
- Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L. Data mining in healthcare and biomedicine: A survey of the literature. *J Med Syst*. 2012;36(4):2431–48. <https://doi.org/10.1007/s10916-011-9710-5>.
- Ferranti D, Krane D, Craft D. The value of prior knowledge in machine learning of complex network systems. *Bioinformatics*. 2017;33(22):3610–8. <https://doi.org/10.1093/bioinformatics/btx438>.
- Witten IH, Frank E, Hall Ma. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed; 2016, p. 664. <http://arxiv.org/abs/arXiv:1011.1669v3>.
- Waegeman W, Dembczyński K, Hüllermeier E. Multi-target prediction: a unifying view on problems and methods. *Data Min Knowl Disc*. 2018;1–32. <https://doi.org/10.1007/s10618-018-0595-5>.
- Kocev D, Vens C, Struyf J, Džeroski S. Tree ensembles for predicting structured outputs. *Pattern Recog*. 2013;46(3):817–33. <https://doi.org/10.1016/j.patcog.2012.09.023>.
- Tsoumakas G, Katakis I. Multi-label classification: An overview. *Int J Data Warehous Min*. 2007;3(3):1–13.
- Zhang ML, Zhou ZH. A review on multi-label learning algorithms. *IEEE Trans Knowl Data Eng*. 2014;26(8):1819–37. <https://doi.org/10.1109/TKDE.2013.39>.
- Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008;4(11):682–90. <https://doi.org/10.1038/nchembio.118>.
- Pei J, Yin N, Ma X, Lai L. Systems Biology Brings New Dimensions for Structure-Based Drug Design. *J Am Chem Soc*. 2014;136(33):11556–65. <https://doi.org/10.1021/ja504810z>.
- Chen X, Yan CC, Zhang XX, Zhang XX, Dai F, Yin J, Zhang Y. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinforma*. 2016;17(4):696–712. <https://doi.org/10.1093/bib/bbv066>.
- Xie L, Xie L, Kinnings SL, Bourne PE. Novel Computational Approaches to Polypharmacology as a Means to Define Responses to Individual Drugs. *Ann Rev Pharmacol Toxicol*. 2012;52(1):361–379. <https://doi.org/10.1146/annurev-pharmtox-010611-134630>.
- Zimmermann GR, Lehár J, Keith CT. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov Today*. 2007;12(1-2):34–42. <https://doi.org/10.1016/j.drudis.2006.11.008>.
- Ding P, Yin R, Luo J, Kwok CK. Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological and Network Knowledge. *IEEE J Biomed Health Inform*. 2018. <https://doi.org/10.1109/JBHI.2018.2852274>.
- Bleakley K, Biau G, Vert J-P. Supervised reconstruction of biological networks with local models. *Bioinformatics* (Oxford, England). 2007;23(13):57–65. <https://doi.org/10.1093/bioinformatics/btm204>.
- Vert J-P, Qiu J, Noble WS. A new pairwise kernel for biological network inference with support vector machines. *BMC Bioinformatics*. 2007;8(Suppl 10):8. <https://doi.org/10.1186/1471-2105-8-S10-58>.
- Vert J-P. Reconstruction of Biological Networks by Supervised Machine Learning Approaches. In: *Elements of Computational Systems Biology*. Hoboken: John Wiley & Sons, Inc.; 2010. p. 163–188. <https://doi.org/10.1002/9780470556757.ch7>. <http://doi.wiley.com/10.1002/9780470556757.ch7>.
- Schrynmackers M, Wehenkel L, Babu MM, Geurts P. Classifying pairs with trees for supervised biological network inference. *Mol BioSyst*. 2015;11(8):2116–25. <https://doi.org/10.1039/c5mb00174a>.
- Pliakos K, Geurts P, Vens C. Global multi-output decision trees for interaction prediction. *Mach Learn*. 2018;107(8-10):1257–81. <https://doi.org/10.1007/s10994-018-5700-x>.
- Yuan Q, Gao J, Wu D, Zhang S, Mamitsuka H, Zhu S. DrugE-Rank: improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*. 2016;32(12):18–27. <https://doi.org/10.1093/bioinformatics/btw244>.
- Bolgár B, Antal P. VB-MK-LMF: fusion of drugs, targets and interactions using variational Bayesian multiple kernel logistic matrix factorization. *BMC Bioinforma*. 2017;18(1):440. <https://doi.org/10.1186/s12859-017-1845-z>.
- Ezzat A, Wu M, Li X-L, Kwok C-K. Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinforma*. 2018. <https://doi.org/10.1093/bib/bby002>.
- Pliakos K, Vens C. Network inference with ensembles of bi-clustering trees. *BMC Bioinforma*. 2019;20(1):525. <https://doi.org/10.1186/s12859-019-3104-y>.
- Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans Comput Biol Bioinforma*. 2004;1(1):24–45. <https://doi.org/10.1109/TCBB.2004.2>.
- Liu Y, Wu M, Miao C, Zhao P, Li X-L. Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. *PLoS Comput Biol*. 2016;12(2):1004760. <https://doi.org/10.1371/journal.pcbi.1004760>.
- Zhang W, Lin W, Zhang D, Wang S, Shi J, Niu Y. Recent Advances in the Machine Learning-Based Drug-Target Interaction Prediction. *Curr Drug Metab*. 2019;20(3):194–202. <https://doi.org/10.2174/1389200219666180821094047>.
- van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*. 2011;27(21):3036–43. <https://doi.org/10.1093/bioinformatics/btr500>.
- Mei J-P, Kwok C-K, Yang P, Li X-L, Zheng J. Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics*. 2013;29(2):238–45. <https://doi.org/10.1093/bioinformatics/bts670>.
- Peng L, Liao B, Zhu W, Li Z, Li K. Predicting Drug-Target Interactions with Multi-Information Fusion. *IEEE J Biomed Health Inform*. 2017;21(2):561–72. <https://doi.org/10.1109/JBHI.2015.2513200>.
- Zong N, Kim H, Ngo V, Harismendy O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics*. 2017;33(15):2337–44. <https://doi.org/10.1093/bioinformatics/btx160>.
- Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, Lu H. Deep-Learning-Based Drug-Target Interaction Prediction. *J Proteome*. 2017;16(4):1401–9. <https://doi.org/10.1021/acs.jproteome.6b00618>.
- Pliakos K, Vens C, Tsoumakas G. Predicting drug-target interactions with multi-label classification and label partitioning. *IEEE/ACM Trans Comput Biol Bioinforma*. 2019. <https://doi.org/10.1109/TCBB.2019.2951378>.

41. Pliakos K. Mining Biomedical Networks Exploiting Structure and Background Information. Belgium: KU Leuven; 2019.
42. Cai R, Zhang Z, Parthasarathy S, Tung AKH, Hao Z, Zhang W. Multi-domain manifold learning for drug-target interaction prediction. In: 16th SIAM International Conference on Data Mining 2016; 2016. p. 18–26. <https://doi.org/10.1137/1.9781611974348.3>.
43. Zhang W, Chen Y, Li D. Drug-target interaction prediction through label propagation with linear neighborhood information. *Molecules*. 2017;22(12): <https://doi.org/10.3390/molecules22122056>.
44. Shi J-Y, Yiu S-M, Li Y, Leung HCM, Chin FYL. Predicting drug-target interaction for new drugs using enhanced similarity measures and super-target clustering. *Methods*. 2015;83:98–104. <https://doi.org/10.1016/j.ymeth.2015.04.036>.
45. Zhang M-L, Zhou Z-H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recog*. 2007;40(7):2038–48. <https://doi.org/10.1016/j.patcog.2006.12.019>.
46. Zhang W, Liu F, Luo L, Zhang J. Predicting drug side effects by multi-label learning and ensemble learning. *BMC Bioinforma*. 2015;16(1): 365. <https://doi.org/10.1186/s12859-015-0774-y>.
47. Ezzat A, Zhao P, Wu M, Li X-L, Kwok C-K. Drug-Target Interaction Prediction with Graph Regularized Matrix Factorization, Vol. 14; 2017. p. 646–56.
48. Ezzat A, Wu M, Li X-L, Kwok C-K. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinforma*. 2016;17(S19): 509. <https://doi.org/10.1186/s12859-016-1377-y>.
49. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics*. 2017;34(7):1164–73. <https://doi.org/10.1093/bioinformatics/btx731>.
50. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. <http://dx.doi.org/10.1023/A:1010933404324>.
51. Li Z-C, Huang M-H, Zhong W-Q, Liu Z-Q, Xie Y, Dai Z, Zou X-Y. Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics*. 2016;32(7):1057–64. <https://doi.org/10.1093/bioinformatics/btv695>.
52. Liu H, Sun J, Guan J, Zheng J, Zhou S. Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics*. 2015;31(12):221–9. <https://doi.org/10.1093/bioinformatics/btv256>.
53. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning - ICML '06. New York: ACM Press; 2006. p. 233–240. <https://doi.org/10.1145/1143844.1143874>.
54. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*. 2015;10(3):0118432. <https://doi.org/10.1371/journal.pone.0118432>.
55. Stock M, Pahikkala T, Airola A., De Baets B, Waegeman W. Efficient Pairwise Learning Using Kernel Ridge Regression: an Exact Two-Step Method. arXiv preprint arXiv:1606.04275. 2016. <http://arxiv.org/abs/1606.04275>.
56. Bekker J, Davis J. Learning From Positive and Unlabeled Data: A Survey. 2018. <http://arxiv.org/abs/1811.04820>.
57. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, Aittokallio T. Toward more realistic drug-target interaction predictions. *Brief Bioinforma*. 2015;16(2):325–37. <https://doi.org/10.1093/bib/bbu010>.
58. Shi J-Y, Li J-X, Chen B-L, Zhang Y. Inferring Interactions between Novel Drugs and Novel Targets via Instance-Neighborhood-Based Models. *Curr Protein Pept Sci*. 2018;19(5):488–97. <https://doi.org/10.2174/1389203718666161108093907>.
59. Shi JY, Zhang AQ, Zhang SW, Mao KT, Yiu SM. A unified solution for different scenarios of predicting drug-target interactions via triple matrix factorization. *BMC Syst Biol*. 2018;12: <https://doi.org/10.1186/s12918-018-0663-x>.
60. Breiman L. Classification and Regression Trees; 1984. <https://doi.org/10.1201/9781315139470>.
61. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3–42. <https://doi.org/10.1007/s10994-006-6226-1>.
62. Van Assche A, Blockeel H. Seeing the Forest Through the Trees: Learning a Comprehensible Model from an Ensemble. In: Machine Learning: ECML 2007. Lecture Notes in Computer Science, Vol 4701. Springer; 2007. p. 418–429. [https://doi.org/10.1007/978-3-540-74958-5\\_39](https://doi.org/10.1007/978-3-540-74958-5_39).
63. Tsoumakas G, Katakis I, Vlahavas I. Mining Multi-label Data. *Data Min Knowl Discov Handbook*. 2010:667–685. [https://doi.org/10.1007/978-0-387-09823-4\\_34](https://doi.org/10.1007/978-0-387-09823-4_34).
64. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*. 2008;24(13):232–40. <https://doi.org/10.1093/bioinformatics/btn162>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

