

RESEARCH

Open Access



TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs

Jian-Yu Shi^{1*}, Hua Huang², Jia-Xin Li¹, Peng Lei³, Yan-Ning Zhang⁴, Kai Dong¹ and Siu-Ming Yiu^{5*}

From 5th International Work-Conference on Bioinformatics and Biomedical Engineering
Granada, Spain. 26-28 April 2017

Abstract

Background: A significant number of adverse drug reactions is caused by unexpected Drug-drug interactions (DDIs). The identification of DDIs becomes crucial before the co-prescription of multiple drugs is made. Such a task in clinics or in drug discovery usually requires high costs and numerous limitations, while computational approaches are able to predict potential DDIs effectively by utilizing diverse drug attributes (e.g. side effects). Nevertheless, they're incapable when required to predict enhancive and degressive DDIs, which change increasingly and decreasingly the pharmacological behavior of interacting drugs respectively. The pharmacological change of DDIs is one of the most important factors when making a multi-drug prescription.

Results: In this work, we design a Triple Matrix Factorization-based Unified Framework (TMFUF) to address the above issue. By leveraging a group of side effect entries of drugs, TMFUF achieves the inspiring result (AUC = 0.842 and AUPR = 0.526) in the case of conventional DDI prediction under the traditional screening task. In the comparison with two state-of-the-art approaches, TMFUF demonstrates its superiority by ~ 7% and ~ 20% improvement in terms of AUC and AUPR respectively. More importantly, TMFUF shows its ability in the comprehensive DDI prediction under different screening tasks. Finally, a utilization TMFUF reveals the significant pairs of side effects, which contribute to form enhancive and degressive DDIs, for further clinical validation.

Conclusions: The proposed TMFUF is first capable to predict both conventional binary DDIs and comprehensive DDIs such that it captures the pharmacological changes caused by DDIs. Furthermore, it provides a unified solution of DDI prediction for two screening scenarios, which involves newly given drugs having no prior interaction. Another advantage is its ability to indicate how significantly the pairs of drug features contribute to form DDIs.

Keywords: Drug-drug interaction, Side effects, Matrix factorization, Prediction, Regression

* Correspondence: jjianyushi@nwpu.edu.cn; smyiu@cs.hku.hk

¹School of Life Sciences, Northwestern Polytechnical University, Xi'an, China

⁵Department of Computer Science, the University of Hong Kong, Hong Kong, China

Full list of author information is available at the end of the article



Background

Two or more drugs in a joint prescription would influence each other in terms of pharmacological behavior [1]. This kind of influence, termed as Drug-Drug Interaction (DDI), could reduce efficacy, induce unexpected toxicities or other adverse drug reactions. Unidentified DDIs would generate unsafe treatments and even medication errors for those patients under the treatment with multi-drug medications [2–5].

Since the number of unidentified DDIs is nearly proportional to the number of newly approved drugs to the power of 2, the broadcasting of DDI-induced adverse effects across medications cannot be negligible. Therefore, DDI identification becomes an urgent need before clinical medications are administered. However, traditionally based on cytochrome P450 testing [6] or transporter-associated interactions [7], the approaches in clinical trials have usually a burden of high cost, long duration and even animal welfare considerations [8], and also face unavoidable challenges, such as inadequate participants and numerous drug pairs to be screened. Consequently, a few DDIs can be identified during drug discovery and development, while most of them are reported in clinics after the corresponding drugs enter the market.

Computational approaches have been developing as a promising assistant of biological/chemical experiments. Both pharmacological research and pharmacy companies pay more attention to them recently [9, 10], because they can rapidly infer potential DDIs among a large number of drug pairs. Current computational approaches can be roughly grouped as text mining-based and machine learning-based approaches. The former detects approved DDIs from diverse text sources [8], such as scientific literature [11, 12], the Adverse Event Reporting System of FDA (<http://www.fda.gov>) and electronic medical records [13]. However, these approaches largely depend on the post-market evidence, such that they cannot alert potential DDIs before multi-drug prescriptions are made. In contrast, the latter can provide such an alert by leveraging the techniques of machine learning (e.g. network recommendation-based [8], Naïve similarity-based approach [14], classification-based [15]). These approaches extract drug features or similarities based on diverse pre-marketed drug properties, such as chemical structures [14], hierarchical classification codes [15], targets [16] and side effects [8]. To the best of our knowledge, the great majority of existing machine learning-based approaches are only able to predict how likely two drugs interact with each other (named as conventional binary prediction). However, two interacting drugs definitely influence each other in terms of pharmacological response *in vivo*.

It is more significant to identify whether DDIs increase or decrease the behaviors of the interacting

drugs in many cases, such as optimizing patient care, establishing drug dosages and finding drug resistance in multi-drug treatments [17]. For two drugs interacting with each other, the occurrence of their interaction would increase or decrease their serum concentration, which is the pharmacological index of measuring the amount of a drug in the pharmacokinetic circulation [18]. For example, when Cyclosporine (who's DrugBank Id is DB00091) is taken with Ticagrelor (DB08816) together, their interaction would increase their serum concentration. While taken with Vincristine (DB00541), the interaction would decrease their serum concentration. Briefly, the first case of DDI is named as an enhancive DDI and the second one is termed as a degressive DDI in the following texts.

In summary, the existing machine learning-based approaches were developed for conventional binary DDIs only, but not for enhancive and degressive DDIs. Moreover, although these approaches can predict the interactions between the drugs having known DDIs and the new drugs having none of existing DDI, they cannot predict the interactions among new drugs. This prediction task is important and helpful to reveal the underlying mechanism of forming DDIs [19].

To address abovementioned issues, this work proposes a Unified Framework of DDI Prediction based on Triple Matrix Factorization (TMFUF). The remaining texts are organized as follows. Section Materials and Method introduces the collection of comprehensive DDIs, the problems of both conventional and comprehensive DDI prediction, the design of our TMFUF, and appropriate cross-validation schemes with respect to different screening scenarios. Section Experiment describes the preparation of DDI prediction, comparison with former approaches in the conventional DDI prediction, and predicting performance of TMFUF under different screening scenarios for both binary and comprehensive prediction. Finally, the last section draws our conclusion.

Methods

Dataset

After searching DrugBank [20], we first acquired 2329 approved small-molecular drugs. Among them, we then removed a set of drugs, which have no DDI entry in DrugBank or have no off-label side effect record in OFF-SIDES [21]. Last, we obtained a set of 603 drugs, of which each drug has at least one DDI and at least one off-label side effect record. Totally, the DDIs among those 603 drugs contains 24,114 DDIs, including 18,710 enhancive DDIs (EnI) and 5404 degressive DDIs (DeI). Moreover, the set of side effects with regard to those 603 drugs finally contains 9149 unique side effect entries, such that each drug can be encoded into a 9149-

dimensional feature vector. See also the next section for technical details.

Moreover, we represent these DDIs as a network, of which nodes are drugs and edges are DDI and then summarize the fundamental properties of the DDI network (Table 1).

Problem formulation of DDI prediction

Denote $\mathbf{D} = \{d_i, i = 1, 2, \dots, m$ as a set of m approved drugs which have known DDIs, and $\mathbf{D}_u = \{d_j, j = 1, 2, \dots, n$ as n newly-given drugs, which have no prior DDIs. The interactions among m approved drugs are organized into an $m \times m$ symmetric interaction matrix $\mathbf{A}_{m \times m} = \{a_{ij}\}$. From the point of view of graph theory, it is the adjacent matrix of a DDI network. In a network of conventional binary DDIs, $a_{ij} \in \{0, 1\}$, where $a_{ij} = 1$ if d_i interacts with d_j , and $a_{ij} = 0$ otherwise. For the comprehensive DDI, $a_{ij} \in \{-1, 0, +1\}$, where $a_{ij} = +1$ indicates an enhancive DDI, $a_{ij} = -1$ indicates a degressive DDI, and $a_{ij} = 0$ indicates no interaction between d_i and d_j respectively. The conventional binary DDI matrix \mathbf{A}_b is just a special case of the comprehensive DDI matrix \mathbf{A} since it can be generated by $\mathbf{A}_b = \text{Binary}(\mathbf{A})$.

Side effects are selected to characterize these drugs, including the approved drugs and the newly-given drugs. According to the clinical occurrence of side effects, each drug d_i is characterized by a high-dimensional feature vector $\mathbf{f}_i = [f_{i1}, f_{i2}, \dots, f_{ik}, \dots, f_{ip}]$, where $f_k = 1$ or 0 indicates whether the k -th specific entry of side effect is observed or not. The feature vectors of the drugs in \mathbf{D} and \mathbf{D}_u are sequentially stacked as an

$m \times p$ feature matrix $\mathbf{F}_{m \times p}$ and an $n \times p$ feature matrix $\mathbf{F}_{n \times p}$ respectively.

For short, a **known/approved drug** is referred to as the drug in \mathbf{D} and a **new drug** is referred to as the drug, which has no known interaction. This paper considers on two screening scenarios involving new drugs (Fig. 1). Their corresponding predicting tasks are defined as follows. The first task (T_1) infers the potential interactions between known drugs and new drugs (e.g. d4 and dx in Fig. 1) while the second one (T_2) infers the potential interactions among new drugs (e.g. dx and dy in Fig. 1). Both of them are required to predict how likely the potential interactions involving new drugs are enhancive or degressive.

Triple matrix factorization and unified predicting model

Since new drugs are isolated nodes in the DDI network (Fig. 1), we cannot deduce their potential interactions by only their topological information. Thus, their additional information (e.g. side effects) is needed. We first extract features based on it and then train a supervised model of DDI prediction.

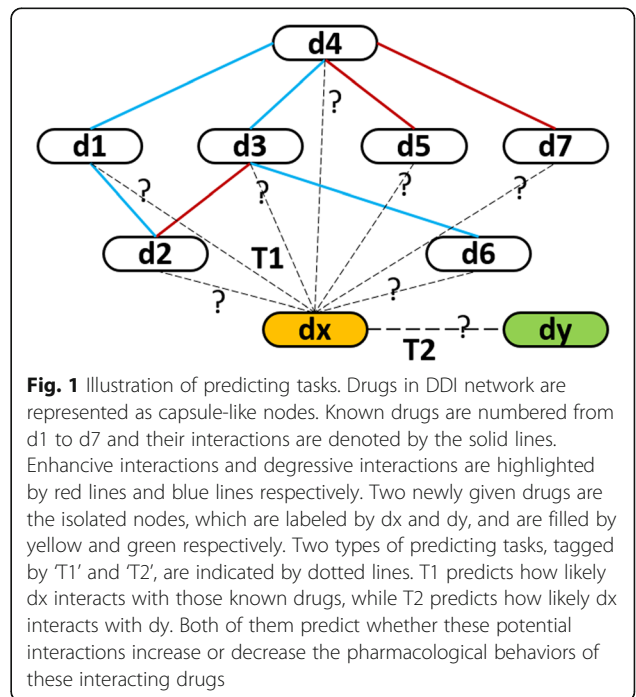
The underlying idea of the proposed model is to connect the features of known drugs in \mathbf{D} with their DDI network topology in a certain way. We model such a connection as a bi-linear regression, which can be represented as a triple matrix factorization (TMF),

$$\mathbf{A} = \mathbf{F}\mathbf{O}\mathbf{F}', \tag{1}$$

where \mathbf{A} is the adjacent matrix of a DDI network, \mathbf{F}

Table 1 Statistics of comprehensive DDI dataset

Property	Entries	Value
Global	Number of Drugs	603
	Number of Interactions	24,114
	Number of Enl	18,710
	Number of Del	5404
Drug Degree	Average Degree of Drug	79.98
	Median Degree of Drug	65
	Max. Degree of Drug	310
	Min. Degree of Drug	1
Drug Enl Degree	Average Enl Degree	62.06
	Median Enl Degree	47
	Max. Enl Degree	242
	Min. Enl Degree	0
	Drug Del Degree	Average Del Degree
Median Del Degree		9
Max Del Degree		219
Min Del Degree		0



denotes the feature matrix of the drugs, and the symmetric projection matrix Θ plays a role of connecting the features of drugs to the interactions between drugs. Each entry θ_{rs} in Θ indicates how well the drug feature pair (f_r, f_s) contributes to form an enhanceive or a degressive interaction. To simplify mathematical symbols in formulas, the apostrophe on “ \mathbf{F} ” in Formula (1) indicates the transpose operator on it. The subsequent formulas in this paper adopt the same notation.

Nevertheless, Formula (1) cannot be solved by $\Theta = (\mathbf{F})^{-1}\mathbf{A}(\mathbf{F}')^{-1}$ directly because of three aspects. First, the inverse matrix of \mathbf{F} doesn't exist when $p \gg m$. Secondly, there exists the multi-collinearity between the columns in \mathbf{F} because the feature entries may not be independent to each other. Thirdly, it is rare to meet an \mathbf{A} of full-rank. Moreover, we may solve it directly by $\Theta^* = \arg \min \|\mathbf{A} - \mathbf{F}_d\Theta\mathbf{F}'_d\|^2$. However, it is very difficult to estimate $p \times p$ entries in Θ when p is large (here $p = 9149$). Thus, we propose a new approach to solve Θ .

We suppose that there exists a low-dimensional latent feature space (interaction space), in which each drug is represented by a latent feature vector and the inner products between drugs are positively correlated with their interactivity. In other words, two drugs possibly form an interaction if they are near to each other and vice versa. Moreover, we also assume that the latent features of the drugs are related to their observed features of drugs. Therefore, the solution of Θ can be achieved as follows

$$\begin{aligned} \mathbf{A}_d^* &= \arg \min \|\mathbf{A} - \mathbf{A}_d\mathbf{A}'_d\|^2 \\ \mathbf{B}^* &= \arg \min \|\mathbf{A}_d^* - \mathbf{F}\mathbf{B}\|^2, \\ \Theta^* &= \mathbf{B}^*(\mathbf{B}^*)' \end{aligned} \tag{2}$$

where the first item in the formula reflects the latent space by matrix factorization, the second one builds the bridge between the observed feature space and the latent feature space of drugs. In detail, \mathbf{A}_d is the $m \times r$ latent interaction matrix, of which each row represents the feature vector of a drug in the latent space. Considering the symmetry of \mathbf{A} , we obtain \mathbf{A}_d by singular value decomposition (SVD) as follows,

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{U}' = \mathbf{U}\sqrt{\Sigma}\sqrt{\Sigma}'\mathbf{U}' = \mathbf{A}_d\mathbf{A}'_d. \tag{3}$$

The $p \times r$ regression coefficient matrix, \mathbf{B} , accounts for the regression between \mathbf{A}_d and \mathbf{F} . The parameter r denotes \mathbf{A} 's rank, which satisfies $r < m$ and reflects the topological complexity of DDI network. So far, we only need to estimate $p \times r$ entries in \mathbf{B} but not $p \times p$ entries in Θ . Partial Least Square Regression is adopted to solve \mathbf{B} , because of $m \ll p$ and the multi-collinearity between feature columns (see also the next section).

After solving Θ^* , we derive the predicting model for both T_1 and T_2 from TMF in a Unified Framework (TMFUF) as follows.

$$\mathbf{A}_{x,D} = \mathbf{F}_x\Theta^*\mathbf{F}', \mathbf{A}_{x,y} = \mathbf{F}_x\Theta^*\mathbf{F}'_y, \tag{4}$$

where \mathbf{F} is the $m \times p$ matrix feature matrix of known drugs, \mathbf{F}_x and \mathbf{F}_y are the $u \times p$ feature matrix of u new drugs $\{d_x\}$ and the $v \times p$ feature matrix of v new drugs $\{d_y\}$ respectively. In addition, $\mathbf{A}_{x,D}$ is the $u \times m$ confidence matrix, of which each entry $a_{x,i}$ indicates how possibly a new drug d_x interacts with a known drug $d_i \in \mathbf{D}$. $\mathbf{A}_{x,y}$ is the $u \times v$ confidence matrix, of which each entry $a_{x,y}$ indicates how possibly d_x interacts with another new drug d_y . The signs of the entries in both of two confidence matrices indicate types of comprehensive DDIs, enhanceive or degressive respectively. Their absolute values are just confidences scores. The larger they are, the more likely they are interactions. Both the training phase and the predicting model are illustrated in Fig. 2.

Partial Least Square regression

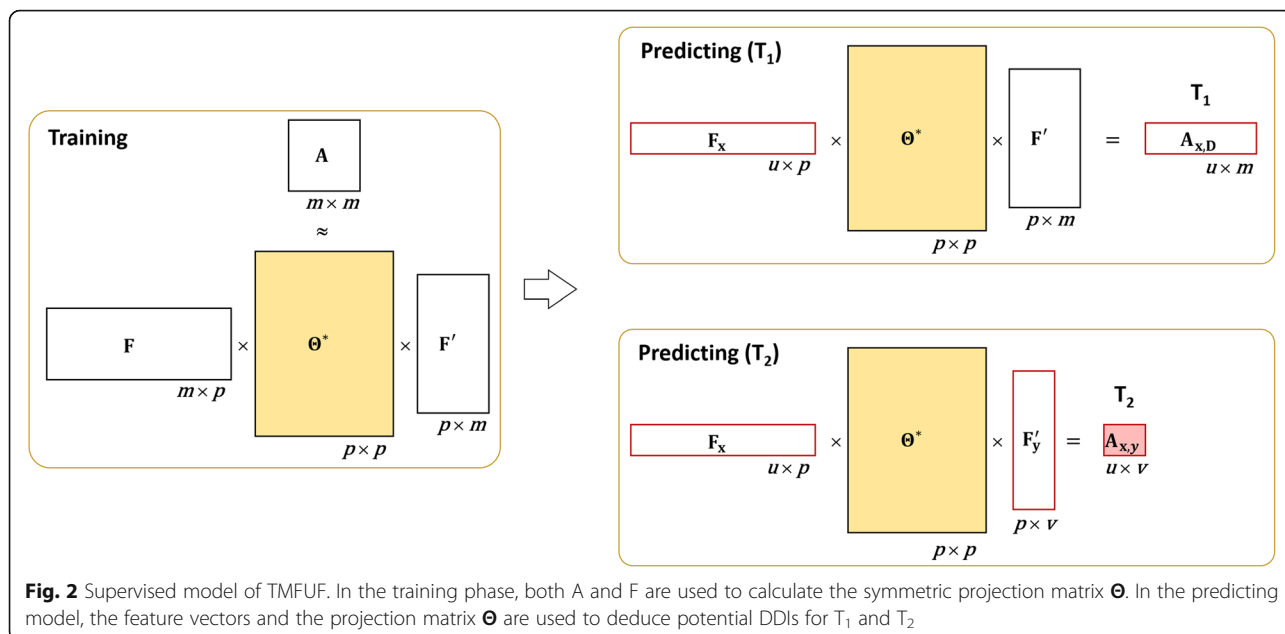
Linear regression can detect the linear relations between two groups of variables (the predictor $\mathbf{X}_{m \times p}$ and the response $\mathbf{Y}_{m \times r}$) with respect to m observations. In this context, we regard drugs as observations, \mathbf{F} as their predictor matrix and \mathbf{A}_d as their response matrix to solve $\mathbf{A}_d \approx \mathbf{F}\mathbf{B}$. See also the second item in Formula 2.

Nevertheless, the standard linear regression fails to solve the linear regression model between \mathbf{F} and \mathbf{A}_d due to the following factors. First, the number of predictors p is significantly greater than that of observations m . Specifically, our dataset contains 603 drugs, of which each is represented as a 9149-dimensional Boolean vector of side effect occurrence ($9149 \gg 603$). Secondly, there is multicollinearity among p columns in \mathbf{F} because some of side effects are highly correlated in terms of the occurrence of side effects (e.g. ‘anaemic hypoxia’ and ‘placental dysplasia’; ‘salivary gland fistula’ and ‘alveolitis necrotising’).

Multivariate Partial Least-Squares Regression (PLSR) is able to overcome this obstacle by mapping both the predictor and the response to a new latent space. Its way to find a linear regression model can be analogue to principal components analysis. Thus, we leverage PLSR to solve the second item in Formula 2. We adopt SIMPLS algorithm to obtain the general underlying model of PLSR since it has only one parameter (i.e. the number of latent factors) to be tuned.

Cross validation

K-fold Cross-validation (K-CV) is one of standard approaches to evaluate the performance of algorithms in machine learning. As former approaches mentioned [19,



22–25], K-CV should be elaborated to avoid over-optimistic results in the case of predicting potential DDIs for new drugs (having no known interaction). Thus, we design two K-CV schemes, CV1 and CV2, when assessing DDI prediction tasks, which are denoted as T_1 and T_2 respectively (see also Fig. 1).

CV1 assesses the prediction that new drugs interact with known drugs (T_1), while CV2 assesses the prediction that some new drugs interact with other new drugs (T_2). Both CV1 and CV2 have several same steps as those in the regular K-CV. The whole dataset of drugs is randomly partitioned into K subsets, of which each contains an approximately equal number of drugs. The drugs in one of K subsets are chosen as the testing drugs and all the drugs in the remaining K-1 subsets are taken as the training drugs.

Remarkably, both CV1 and CV2 regard drug pairs but not drugs as samples. They have in common is that the drug pairs consisting of only the training drugs are selected as the training samples, while their difference is the selection of the testing samples.

- In each round of CV1, the testing samples are the drug pairs between the testing drugs and the training drugs, because their labels should be blind to the training phase. Such a selection of the testing samples reflects that the testing drugs (imitating new drugs) have no prior interaction with any of the training drugs. CV1 repeats the training and the testing K times by taking the drugs in each subset as the testing drugs in turn, and it averages the predicting performances in K rounds as its final performance.
- In each round of CV2, the testing samples are only the drug pairs among the testing drugs and their

labels are blind to the training phase as well. Especially, the drug pairs between the testing drugs and the training drugs are rejected in the training phase, because they contain the information about the testing samples in CV2 and also discarded in the testing phase because they are the testing samples in CV1. Each round of CV2 take two subsets of drugs to label the testing drug pairs at one time. When these two subsets are same, these are K cases, which are complementary to those of CV1. When they are different to each other, there are $K \times (K-1)/2$ cases. Totally, CV2 contains $K + K \times (K-1)/2 = K \times (K+1)/2$ rounds and its final performance is also the average of the predicting performance of these rounds.

To illustrate both CV1 and CV2, we show a toy case containing nine drugs (Fig. 3). To perform the cross validation, they're first randomly shuffled and renumbered in ascending order. Then we perform the sampling for 3-fold cross validation sequentially.

In CV1 (Fig. 3a), we split the drugs into three exclusive groups, of which each is sampled as the testing drugs and the remaining two groups are sampled as the training drugs in turn. Consequently, CV1 performs $C_K^1 = K = 3$ rounds of CV. For example, in the second round of CV1, there are six training drugs (labeled as d_1, d_2, d_3, d_7, d_8 and d_9) and three testing drugs (labeled as d_4, d_5 and d_6). A similar separation of training and testing drugs can be found in two other rounds of CV1. In Fig. 3a, the training interactions among the training drugs are denoted by white cells and used to build the predicting model, while

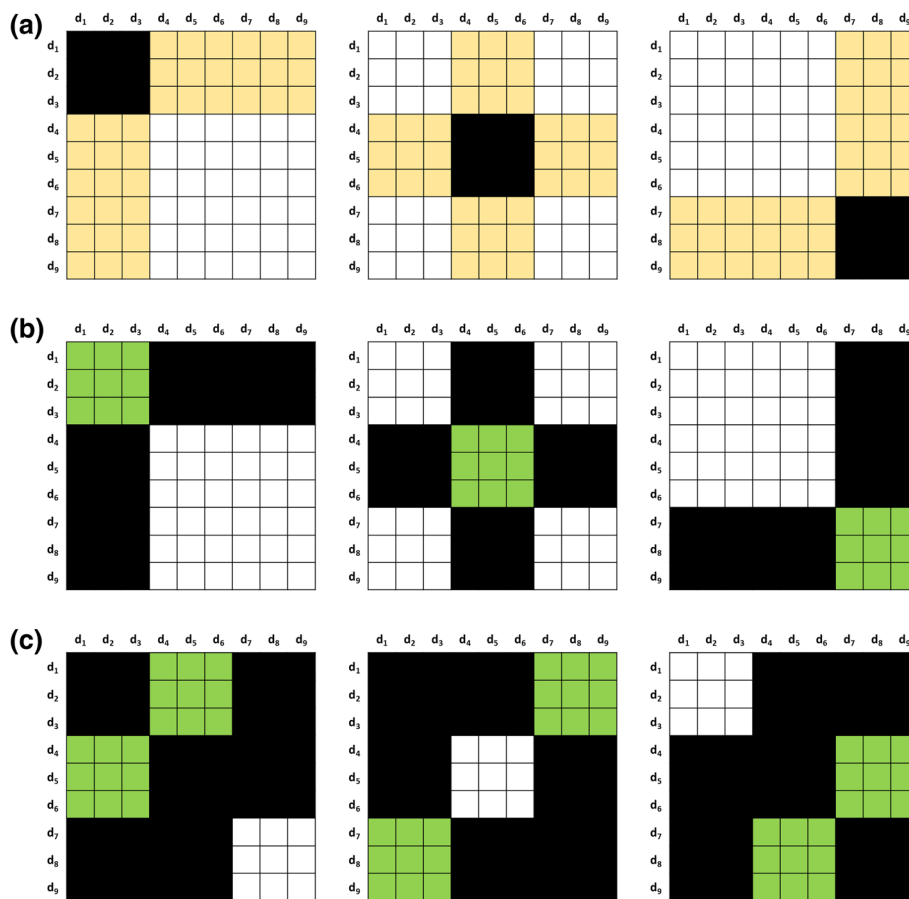


Fig. 3 Illustration of K-CV schemes. The 9×9 DDI adjacent matrix is represented by a table with 9×9 cells. It is split into different blocks, which account for training, testing and discard parts, and are filled with different colors accordingly. All the entries in the white block are used to train the TMFUF, the entries in the yellow and the green block denote the testing entries in T_1 and T_2 respectively, and the entries in the black region are blind to the training and testing. Three subfigures show different 3-CV schemes: **a** the sampling in CV1, **b** the within-group sampling in CV2, and **c** the between-group sampling in CV2

the testing interactions between the set of the testing drugs and the set of the training drugs are denoted by yellow cells in DDI adjacent matrix. Black cells are blind to both training and testing in CV1. Totally, CV1 contains three rounds of testing.

Being stricter than CV1, CV2 contains two kinds of drug sampling, named within-group and between-group samplings. In the within-group sampling of CV2 (Fig. 3b), we also split the drugs into three groups, of which each is sampled as the testing drugs and the other two groups are sampled as the training drugs in turn. Likewise, the within-group sampling performs $C_K^1 = K = 3$ rounds of CV. Until here, the procedure is exactly same as that of CV1. However, their difference is what the testing interactions are, because this sampling simulates the scenario of predicting DDI among a set of newly given drugs. Thus, the within-group sampling takes the interactions within the testing group of drugs as the

testing interactions (green cells in Fig. 3b), and the interactions among the training drugs as the training interactions (white cells in Fig. 3b). Meanwhile, the interactions between the set of the training drugs and the set of the testing drugs (black cells in Fig. 3b) are blind to both training and testing in the within-group sampling. Obviously, the within-group sampling of CV2 can be regarded as the complement of the sampling of CV1.

The between-group sampling provides a stricter way than the within-group sampling. It simulates the scenario of predicting DDI between two different sets of newly given drugs, of which any has no known interactions. In each round of the between-group sampling, two different groups are selected as the testing set and the other group of drugs is selected as the training set. Consequently, because of the symmetry of DDI adjacent matrix, the between-group sampling performs $C_K^2 = K(K-1)/2 = 3$ rounds of CV. In the round, the testing

interactions are only those between two testing groups of drugs (green cells in Fig. 3c) and the training interactions are only those among the training drugs (white cells in Fig. 3c). Remarkably, the interactions between the drugs in any of the testing groups and the training drugs (black cells in Fig. 3c) and the interactions among the drugs in any of the testing groups (black cells in Fig. 3c) are blind to both training and testing in the between-group sampling of CV2. Therefore, CV2 contains $K \times (K + 1) / 2 = 6$ rounds of testing in total.

To sum up, CV1 accounts for the screening scenario that we're asked to infer how possibly a newly given drug interacts with one or more known drugs. It is particularly useful before we enlarge existing multi-drug prescriptions by appending new drugs.

While CV2 accounts for another screening scenario that we are required to determine how possibly a newly-given drug interacts with another new one. It is definitely helpful when the need to make novel multi-drug prescriptions arises. In addition, it is one of the most crucial steps towards understanding and revealing how DDIs form.

Receiver operating characteristic (ROC) curve and precision-recall (PR) curve are two of popular approaches to illustrate the performance of computational approaches in binary classification [26]. In addition, both the area under ROC curve (AUC) and the area under PR curve (AUPR) are usually adopted to measure the performance of binary classification or prediction [26].

The prediction of conventional DDIs is a typical binary prediction, in which interactions and non-interaction are labeled as positive and negative samples respectively. Thus, we can directly draw AUC and PR curves for the prediction and measure its performance with both AUC and AUPR by comparing the predicted scores of positive samples with those of negative samples.

To draw ROC and PR curves for the prediction of comprehensive DDIs, we reverse the labels of degressive DDIs and their predicting scores. Degressive DDIs are first labeled as positive samples. Then, their new scores are generated by the minus of their original predicted scores. After that, united with enhance DDIs, they are treated as positive samples. Last, the same procedure as that of conventional DDIs is adopted to measure the prediction performance of comprehensive DDIs.

Results and discussion

Preparation and parameter tuning

When performing when performing CV1 for T_1 or CV2 for T_2 , we set $K = 10$ such that CV1 and CV2 contain 10 and 55 rounds of the training-testing phases respectively.

To achieve a statistical significance, we generated the partition of the training of the testing drugs under 50 different random seeds, and repeated CV1 and CV2 50 times accordingly. We reported the average performance over 50 repetitions of CV as the final evaluation of the prediction.

The number of latent factors (denoted as L in PLSR) is the only one tunable parameter in TMFUF. To obtain its best value, we adopted a simple way that performs a series of binary DDI predictions under 10CV in task T_1 along with different values of L and measured their performance by AUC. A fixed list $\{1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150\}$ was given for tuning L . Different values of L give different predicting results. The value ($L = 60$) corresponding to the maximum AUC was selected out as the best value of L (Fig. 4) and further applied in the subsequent experiments.

Comparison with state-of-the-art

In order to validate the effectiveness of our TMFUF, we first compared it with two recent works, Naïve similarity-based approach [14] and label propagation-based approach [8], in the scenario of T_1 with respect to conventional binary DDIs. The former approach infers how likely a newly given drug interacts with one known drug of interest by the summation of its similarities to those drugs interacting with the same drug of interest as well. The latter approach directly leverages Label Propagation (a semi-supervised classifier) to infer potential DDIs by regarding the binary DDI network as a binary label network.

In the binary prediction of DDIs, we generated the adjacent matrix of binary DDIs by simply turning “-1”s into “+1” in that of comprehensive DDIs. The results demonstrate that TMFUF exceeds these two state-of-the-art approaches significantly with $\sim 7\%$ and $\sim 20\%$ improvements of AUC and AUPR respectively (Table 2).

Furthermore, we noted that these approaches were only designed for the conventional prediction of binary DDIs in the scenario of T_1 . Therefore, we also demonstrated the ability of TMFUF over four predicting scenarios with respect to two pharmacological changes caused by DDI and two screening needs for multi-drug treatments. The results are listed in Table 3 and the corresponding ROC curves are shown in Fig. 5. It can be observed that the performance of comprehensive prediction is worse than that of conventional binary prediction. For example, in task T_1 , the comprehensive prediction shows 73.3% AUC and 48.3% AUPR, while the conventional prediction shows 84.2% and 52.6% respectively. In addition, it is found that T_2 is a more difficult task than T_1 . Compared with T_1 , T_2 generally exhibits $\sim 15\%$ and $\sim 22\%$ degradation in terms of both AUC and AUPR.

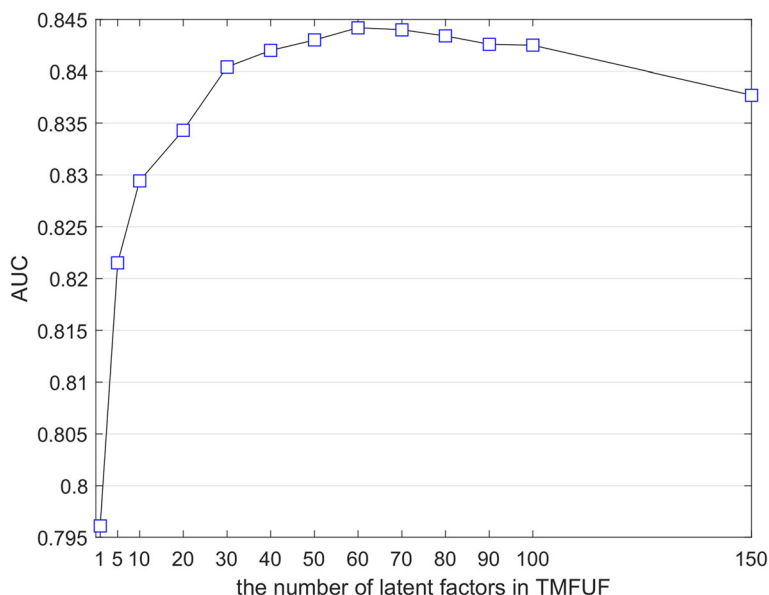


Fig. 4 Illustration of Parameter Tuning. The number of the latent factors is tuned by a fixed list {1,5,10,20,30,40,50,60,70,80,90,100,150}. The final number is determined by the value corresponding to the maximum AUC

Significant feature pairs

According to Formula (1), we achieved the symmetric projection matrix Θ , which bridges drugs’ features and their interactions with the trivial reconstructed error 29.65 in terms of Frobenius Norm among 181,503 drug pairs. Especially, Θ is able to provide the indication of significant pairs of side effect entries. Because the number of feature entries is large, the values of the entries in Θ would be very small. To avoid the computational errors, we multiplied A by 1000 before solving Θ . The histogram of the entry values shows that very most of them are near to ZERO (Fig. 6). Therefore, we defined the entries, whose absolute values are greater than 1, as Significant Entries.

The subscripts of those significant entries correspond to the indices of features. Positive and negative significant entries indicate enhancive DDIs and degressive DDIs respectively. After sorting the entries by their absolute values, we obtained a set of significant pairs of side effects. Top 10 significant pairs for enhancive DDIs and degressive DDIs are listed in Tables 4 and 5 respectively.

Table 2 Comparison with state-of-the-art methods in the scenario of T1 with respect to binary DDIs

Method	AUC	AUPR
Naïve Similarity [14]	0.759 ± 0.001	0.302 ± 0.002
Label Propagation [8]	0.774 ± 0.001	0.326 ± 0.002
TMFUF	0.842 ± 0.002	0.526 ± 0.006

The best results are highlighted in boldface

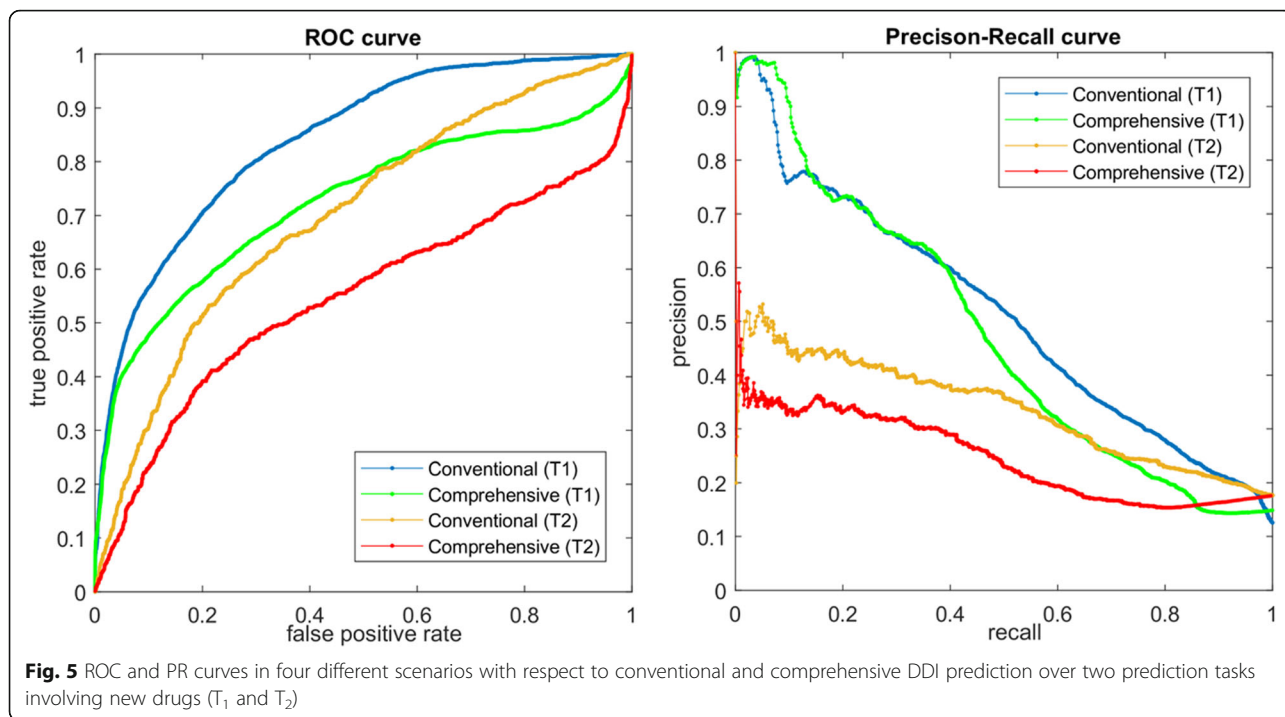
There are several observations about these significant side effects. According to the frequency of their occurrences, the dose-related side effects and the cardiovascular system-related side effects are two types of the most significant factors in overall DDIs. The former includes ‘overdose’, ‘multiple drug overdose’ and ‘dependence’. The latter includes ‘accelerated hypertension’, ‘blood sodium’, and ‘bradycardia’. Especially, ‘overdose’ is dominant in enhancive DDIs while ‘accelerated hypertension’ is dominant in degressive DDIs. Thus, we believe that both the control of safe dosage and the adverse effects on the cardiovascular system are two important factors, which should be considered when screening DDIs in the clinic.

Conclusions

There is a need to screen DDIs before multi-drugs prescriptions are made. Current computational approaches can rapidly deduce potential DDI candidates among a large number of drug pairs with a low cost. Nevertheless, they have two weak points. First, they are just appropriate for predicting binary

Table 3 Predicting Performance of TMFUF under Different Scenarios

Task	AUC		AUPR	
	Conventional	Comprehensive	Conventional	Comprehensive
T1	0.842 ± 0.002	0.733 ± 0.004	0.526 ± 0.006	0.483 ± 0.007
T2	0.702 ± 0.004	0.577 ± 0.005	0.303 ± 0.005	0.246 ± 0.005



DDIs, but incapable for predicting comprehensive DDIs, which increase or decrease the pharmacological behaviors of the interacting drugs in vivo. Secondly, although they are able to predict whether a new drug (having no approved interactions) interact with known drugs (have approved interactions), but cannot predict whether or not a new drug interacts with another.

To address the issues about DDIs types and screening scenarios, we have proposed a novel approaches (TMFUF) for predicting DDIs. TMFUF presents a unified solution for the prediction of both conventional binary DDIs and comprehensive DDIs under different screening scenarios. It is not only useful before we enlarge existing multi-drug prescriptions by appending new drugs, but also helpful when we need to make novel multi-drug prescriptions consisting of only new drugs. The superiority of TMFUF is demonstrated in the prediction of conventional binary and comprehensive DDIs under two screening scenarios respectively. What's more, TMFUF is able to uncover

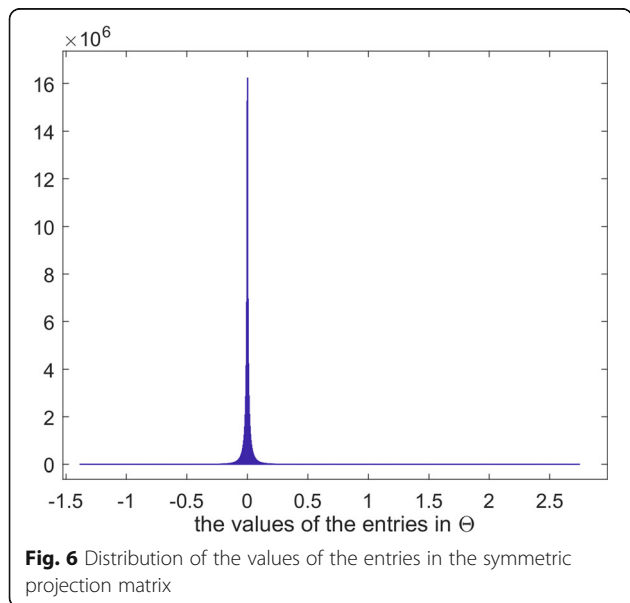


Table 4 Top 10 Significant Pairs for Enhance DDIs

Rank	Side Effect 1	Side Effect 2
1	'multiple drug overdose'	'overdose'
2	'blood sodium'	'bradycardia'
3	'blood sodium'	'overdose'
4	'accelerated hypertension'	'diarrhea'
5	'glucose intolerance'	'overdose'
6	'diabetic neuropathy'	'overdose'
7	'dependence'	'overdose'
8	'blood sodium'	'multiple drug overdose'
9	'bradycardia'	'overdose'
10	'blood sodium'	'drug interaction'

Table 5 Top 10 Significant Pairs for Degressive DDIs

Rank	Side Effect 1	Side Effect 2
1	'accelerated hypertension'	'hypotonia'
2	'burning sensation mucosal'	'overdose'
3	'accelerated hypertension'	'overdose'
4	'accelerated hypertension'	'convulsion'
5	'accelerated hypertension'	'diabetes mellitus inadequate control'
6	'accelerated hypertension'	'neuroleptic malignant syndrome'
7	'accelerated hypertension'	'dependence'
8	'accelerated hypertension'	'cholecystitis chronic'
9	'gastrointestinal disorder'	'overdose'
10	'abuse'	'accelerated hypertension'

significant feature pairs, which contribute to enhance and degressive DDIs. In the future, we will integrate other features, (e.g. chemical structures, drug-binding proteins) into the framework of TMFUF to achieve better DDI prediction, such that it is hopeful to reveal the underlying mechanism of how DDIs form.

Abbreviations

AUC: The area under the receiver operating characteristic curve; AUPR: The area under precision-recall curve; CV: Cross-validation; DDI: Drug-drug interaction; PLSR: Partial Least-Squares Regression; TMFUF: Triple Matrix Factorization-based Unified Framework

Acknowledgements

The preliminary version of this work was previously published in the Proceedings of the 5th International Work-Conference on Bioinformatics and Biomedical Engineering, Lecture Notes in Computer Science: Bioinformatics and Biomedical Engineering [27].

Funding

This work was supported by RGC Collaborative Research Fund (CRF) of Hong Kong (C1008-16G), National High Technology Research and Development Program of China (No. 2015AA016008), the Fundamental Research Funds for the Central Universities of China (No. 3102015ZY081), the Program of Peak Experience of NWPU (2016), China National Training Programs of Innovation and Entrepreneurship for Undergraduates (No. 201710699330), and the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University (No. ZZ2018170, ZZ2018235). The publication charge was funded by China National Training Programs of Innovation and Entrepreneurship for Undergraduates (No. 201710699330) and the National Natural Science Foundation of China (No. 61872297).

Availability of data and materials

The dataset used in this paper can be download from <https://github.com/JustinShi2016/IWBIO2017>

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 19 Supplement 14, 2018: Selected articles from the 5th International Work-Conference on Bioinformatics and Biomedical Engineering: bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-14>.

Authors' contributions

JYS conceived the computational methodology, designed the experiments and draft the manuscript. Both HH and JXL collected DDI entries from

DrugBank and side effect entries from OFFSIDES, and performed the experiments. JYS, HH, JXL developed the codes of both the proposed approach and the analysis of results jointly. PL and KD analyzed the results, especially on significant feature pairs. Both YNZ and SMY also helped draft the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹School of Life Sciences, Northwestern Polytechnical University, Xi'an, China. ²School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China. ³Department of Chinese Medicine, Shaanxi Provincial People's Hospital, Xi'an, China. ⁴School of Computer Science, Northwestern Polytechnical University, Xi'an, China. ⁵Department of Computer Science, the University of Hong Kong, Hong Kong, China.

Published: 20 November 2018

References

- Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov*. 2005;4(10):825–33.
- Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, Hallisey R, Ives J, Laird N, Laffel G, et al. Systems analysis of adverse drug events. ADE prevention study group. *JAMA*. 1995;274(1):35–43.
- Businaro R. Why we need an efficient and careful pharmacovigilance. *Aust J Pharm*. 2013;1:4.
- Karbownik A, Szalek E, Sobańska K, Grabowski T, Wolc A, Grześkowiak E. Pharmacokinetic drug-drug interaction between erlotinib and paracetamol: a potential risk for clinical practice. *Eur J Pharm Sci*. 2017;102:55–62.
- Mulroy E, Highton J, Jordan S. Giant cell arteritis treatment failure resulting from probable steroid/antiepileptic drug-drug interaction. *N Z Med J*. 2017; 130(1450):102–4.
- Veith H, Southall N, Huang R, James T, Fayne D, Artemenko N, Shen M, Inglesse J, Austin CP, Lloyd DG, et al. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat Biotechnol*. 2009;27(11):1050–5.
- Huang SM, Temple R, Throckmorton DC, Lesko LJ. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther*. 2007;81(2):298–304.
- Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep*. 2015;5:12339.
- Wiśniowska B, Polak S. The role of interaction model in simulation of drug interactions and QT prolongation. *Curr Pharmacol Rep*. 2016;2(6):339–44.
- Zhou D, Bui K, Sostek M, Al-Huniti N. Simulation and prediction of the drug-drug interaction potential of Naloxegol by physiologically based pharmacokinetic modeling. *CPT: Pharmacometrics Syst Pharmacol*. 2016; 5(5):250–7.
- Bui QC, Sloat PMA, van Mulligen EM, Kors JA. A novel feature-based approach to extract drug-drug interactions from biomedical text. *Bioinformatics*. 2014;30(23):3365–71.
- Zhang Y, Wu HY, Xu J, Wang J, Soysal E, Li L, Xu H. Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug drug interactions from biomedical literature. *BMC Syst Biol*. 2016;10(Suppl 3):67.
- Duke JD, Han X, Wang ZP, Subhadarshini A, Karnik SD, Li XC, Hall SD, Jin Y, Callaghan JT, Overhage MJ, et al. Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interactions. *PLoS Comput Biol*. 2012;8(8): e1002614.

14. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, Tatonetti NP. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc.* 2014;9(9):2147–63.
15. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *JAMIA: J Am Med Inform Assoc.* 2014;21(e2):e278–86.
16. Luo H, Zhang P, Huang H, Huang J, Kao E, Shi L, He L, Yang L. DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome. *Nucleic Acids Res.* 2014;42(Web Server issue):46–52.
17. Koch-Weser J. Serum drug concentrations in clinical perspective. *Ther Drug Monit.* 1981;3(1):3–16.
18. Bichteler A, Wikoff DS, Loko F, Harris MA. Estimating serum concentrations of dioxin-like compounds in the US population effective 2005-2006 and 2007-2008: a multiple imputation and trending approach incorporating NHANES pooled sample data. *Environ Int.* 2017;105:112–25.
19. Shi JY, Li JX, Gao K, Lei P, Yiu SM. Predicting combinative drug pairs towards realistic screening via integrating heterogeneous features. *BMC Bioinformatics.* 2017;18(Suppl 12):409.
20. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014;42(Database issue):D1091–7.
21. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med.* 2012;4(125):125ra131.
22. Pahikkala T, Airoola A, Pietila S, Shakyawar S, Sz wajda A, Tang J, Aittokallio T. Toward more realistic drug-target interaction predictions. *Brief Bioinform.* 2015;16(2):325–37.
23. Shi JY, Yiu SM, Li YM, Leung HCM, Chin FYL. Predicting drug-target interaction for new drugs using enhanced similarity measures and super-target clustering. *Methods.* 2015;83:98–104.
24. Shi J-Y, Gao K, Shang X-Q, Yiu S-M. LCM-DS: a novel approach of predicting drug-drug interactions for new drugs via Dempster-Shafer theory of evidence. In: *BIBM. IEEE;* 2016. p. 512–5.
25. Shi J-Y, Li J-X, Lu H-M. Predicting existing targets for new drugs base on strategies for missing interactions. *Bmc Bioinformatics.* 2016;17(Suppl 8):282.
26. Jiao Y, Du P. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant Biol.* 2016;4(4):320–30.
27. Shi J-Y, Huang H, Li J-X, Lei P, Zhang Y-N, Yiu S-M: Predicting comprehensive drug-drug interactions for new drugs via triple matrix factorization. In: *WWWBIO: 2017; Spain. Lecture Notes in Computer Science: Bioinformatics and Biomedical Engineering, Springer;* 2017: 108–117.

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

