Predicting drug-drug interactions based on integrated similarity and semi-supervised learning

Cheng Yan, Guihua Duan*, Yayan Zhang, Fang-Xiang Wu, Yi Pan and Jianxin Wang

Abstract—A drug-drug interaction (DDI) is defined as an association between two drugs where the pharmacological effects of a drug are influenced by another drug. Positive DDIs can usually improve the therapeutic effects of patients, but negative DDIs cause the major cause of adverse drug reactions and even result in the drug withdrawal from the market and the patient death. Therefore, identifying DDIs has become a key component of the drug development and disease treatment. In this study, we propose a novel method to predict DDIs based on the integrated similarity and semi-supervised learning (DDI-IS-SL). DDI-IS-SL integrates the drug chemical, biological and phenotype data to calculate the feature similarity of drugs with the cosine similarity method. The Gaussian Interaction Profile kernel similarity of drugs is also calculated based on known DDIs. A semi-supervised learning method (the Regularized Least Squares classifier) is used to calculate the interaction possibility scores of drug-drug pairs. In terms of the 5-fold cross validation, 10-fold cross validation and de novo drug validation, DDI-IS-SL can achieve the better prediction performance than other comparative methods. In addition, the average computation time of DDI-IS-SL is shorter than that of other comparative methods. Finally, case studies further demonstrate the performance of DDI-IS-SL in practical applications.

Index Terms—Drug,Drug-drug interactions,Regularized least squares classifier,Gaussian interaction profile kernel similarity .

1 Introduction

The pharmacological effect of a drug is influenced by another drug, which usually appears when two or more drugs are administered simultaneously for a patient. These associations are also defined as drug-drug interactions (DDIs), and are either favorable efficacy or undesirable DDIs according to clinical results. Positive DDIs can provide more effective treatments and reduce the suffering of patients. However, undesirable DDIs are the major cause of adverse reaction events [1]. In serious cases, they can result in the drug withdrawal from the drug market and the death of a patient who is treated with multi-drugs [2], [3]. Currently, multi-drug therapies have been widely used in treating multiple illnesses or complex diseases, such as cancer [4], [5], [6]. The original purpose of multi-drugs treatment is to alleviate the patient suffering, improve the treatment effect and increase the overall survival rate [7]. However, undesirable DDIs have also been developed along with more and more drugs used in the synergistic treatment, and which also influence the treatment effect and even lead to serious complications as well as the financial burden. Therefore, in order to reduce the cost of drug development and improve the treatment effect, it is very urgent to identify DDIs in the drug development process.

Recently, many studies have proven that some commonly used drugs have high possibility to interact with each other, such as lipidlowering drugs, macrolides, oral antifungal agents, which are widely used to synergistic treatments [8], [9], [10]. Previous studies about DDIs can be divided into three categories: pharmaceutic, pharmacokinetic (PK) and pharmacodynamic (PD) [11], [12]. The pharmaceutic DDIs usually result from multi-drugs with the chemical incompatibility. A PK interaction is defined as the effects of a drug in the absorbed, distributed, or metabolized process of another drug in the patient body, which is usually related to adverse responses [4]. PD interactions often result from different drugs acting on the same receptor, site, or physiological system, and could have also either synergistic or harmful effects for patients [13]. Many PK and PD interactions have been used for inferring DDIs in previous studies [14], [15].

In silico, in vitro and in vivo experiments are the methods to discover DDIs among drugs, and the two latter methods are usually very time-consuming and labor-intensive cycles [16]. In addition, the side effects caused by DDIs are hard to be measured in vitro or in vivo experiments, which makes results that these methods hard to be executed [16]. As more and more patients are simultaneously treated by multi-drugs, identifying DDIs has become an important issue of bioinformatics research and a very urgent need to drug developments. Moreover, compared to traditional biomedical experiment methods, the computational methods provide an opportunity to predict new DDIs with the low cost and high accuracy. Therefore, by considering its

- *C.Yan, G.Duan, Y.Zhang and J.Wang are with the School of Computer Science and Engineering, Central South University, Changsha 410083, China.*
  E-mail: {yancheng01, duangh, 174611156, jxwang}@mail.csu.edu.cn
- *C.Yan is also with the School of Computer and Information, Qiannan Normal University for Nationalities,558000, Duyun, Guizhou, China*
  E-mail: yancheng01@mail.csu.edu.cn
- *FX.Wu is with the Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SKS7N5A9, Canada.*
  E-mail: faw341@mail.usask.ca
- *Y.Pan is with the Department of Computer Science, Georgia State University, Atlanta, GA30302, USA.*
  E-mail: yipan@gsu.edu
- ** Corresponding author*

advancement to biological experiments, there exists a high demand for predicting DDIs via computational approaches [17]. In addition, the development of medical technologies and applications of multi-drug treatments also further imposes a very urgent demand to develop computational methods to predict potential DDIs.

Recently, based on machine learning models, many computational approaches have been developed to predict potential DDIs. Tatonetti *et al*. developed a signal discovery method to infer DDIs [18], main features of drugs used in this method are drug adverse event profiles. By combining drug chemical similarities, side effect similarities, protein-protein interaction similarities and target sequence similarities, an INDI (INferring Drug Interactions) framework was developed to predict DDIs, which used two types of drug interactions(potential CYP (Cytochrome P450)-related DDIs, and non-CYP-related DDIs (NCRDs)) [19]. By the combination of crizotinib with ketoconazole or rifampin, a PBPK (physiologically based pharmacokinetic) model was developed for predicting DDIs [20]. Based on properties of the drug metabolism, the text-mining and reasoning approaches were also used to discover novel DDIs [21]. Vilar *et al*. computed the molecular fingerprint similarity and the molecular structure similarity of drugs to predict DDIs [22]. With 2D and 3D molecular structures, interaction profiles, target and side-effect similarities, Vilar *et al*. further developed a protocol applicable on a large scale data to infer novel DDIs [23]. Based on drug phenotypic, therapeutic, chemical, and genomic properties and machine learning model, Cheng *et al*. proposed a computational method to predict DDIs [24]. Based on the drug molecular similarity and phenotypic similarity, Li *et al*. developed a computational method to discover the combination efficacy of drugs with a Bayesian network model [25]. Based on a random forest model, Liu *et al*. proposed a computational method to predict DDIs by integrating chemical interactions, protein-protein interactions between targets of drugs and target enrichment of KEGG pathways [26]. This method adopted a feature selection technique to obtain the important features of drugs. Luo *et al*. developed a computational method to predict DDIs by implementing the chemical-protein interactome, which provided as a web server (called DDI-CPI) [27]. Based on the framework probabilistic soft logic, Sridhar *et al*. took a PSL (Probabilistic Soft Logic) method to predict novel DDIs by integrating networks of multiple drug similarities and known DDIs [13]. With 2D structural similarities of drugs, Takako *et al*. developed a logistic regression model to infer potential DDIs [28]. Its prediction performance is further improved by combining target-related and enzyme-related scores. Based on inner product-based similarity measures (IPSMs), Ferdousi *et al*. provided an computational method to predict DDIs. This method also used the drug similarity constructed with key biological elements including carriers, transporters, enzymes and targets of drugs. In addition, based on the assumption that synergistic effects with drugs are often similar and vice versa, NLLSS (Network-based Laplacian regularized Least Square Synergistic drug combination prediction) was proposed to predict hidden synergistic drug combinations, but it can not predict DDIs for new drugs [29].

In addition, based on the network-based prediction models, many prediction methods have been developed for the DDI prediction and other closely related issues, such as the drug-disease prediction and the drug-target interaction prediction [30], [31]. The network-based DDI prediction methods can be divided into two types: one is constructing a similarity network of known DDIs to predict DDIs while another is performing the novel DDI prediction based on the structure of the DDI network. For example, the MEF (multiple evidence fusion) method was developed for predicting adverse drug reactions (ADRs), which is based on the network structural data of similarities of drugs and known drug-ADR interactions [32]. Cao *et al*. also provided a similarity network-based method to predict hidden DDIs by computed relational features of drugs [24]. The label propagation method is a typical network-based method, which is also used to infer new DDIs by integrating drug chemical structures, drug side effects and off side effects [33]. Based on the drug-target interaction and protein-protein interaction networks, potential PK DDIs were predicted by analyzing the significant relation between ”S-score” and likelihood that DDIs occurs [34]. By computing the drug similarity via integrating the drug-target interactions and protein-protein interactions, Park *et al*. adopted a random walk with restart model to calculate the DDI scores which was the likelihood of the occurrence of PD DDIs [35]. The DDI type is also an important issue of the DDI prediction, Jin *et al*. provided a computational method to infer the DDI type by applying a multi-task dyadic regression model [36]. Based on the Jaccard similarity index and interaction profile fingerprints (IPFs), Vilar *et al*. developed a computational method to predict new DDIs [37]. Furthermore, by integrating the chemical, biological, phenotype and known DDI network information, Zhang *et al*. developed an effective approach to predict potential DDIs, it explored three ensemble methods which include the weight average ensemble method, L1 ensemble (L1E) classifier and L2 ensemble (L2E) classifier.

However, although those above computational methods have achieved some effective results for predicting novel DDIs, some limits still should be addressed. For example, some methods have not effectively integrated the known DDI network information and features of drugs by a reasonable model. Although there exist a large number of new drugs, these current computational methods have not paid enough attention to discover potential DDIs for them. With the development of multi-drug treatments, the possibility of adverse effect occurrences is higher than before. Therefore, in order to improve the effect of multi-drug treatments and drug developments, it is very urgent to develop more effective computational methods to predict potential DDIs.

In this study, by integrating the chemical, biological and phenotype information of drugs, we develop a computational method (called DDI-IS-SL) to predict DDIs. These drug information includes drug chemical structures, drug-target interactions, drug enzymes, drug transports, drug pathways, drug indications, drug side effects, drug off side effects and known DDIs. First, based on these pieces of drug information, we construct a high-dimensional binary vector to calculate the feature similarity of drugs via the cosine similarity method. Furthermore, we also compute the Gaussian Interaction Profile (GIP) kernel similarity [38] of

drugs based on known DDIs. The final drug similarity is constructed by their feature similarity and GIP similarity. Then a Regularized Least Squares (RLS) classifier [39] is adapted to predict DDIs. For new drugs which do not have any interactions with other drugs, we also calculate their relational initial scores via performing the node-based drug network diffusion method. Therefore, our method can predict potential DDIs not only for known drugs but also for new drugs. The prediction performance of our method and other competing methods are systematically assessed by the 5-fold cross validation, the 10-fold cross validation and the de novo validation. The AUC (area under the ROC curve) is used as the metric to evaluate the performance of computational methods. In terms of AUC, our method is superior to other competing methods. Specifically, in the 5-fold cross validation, the AUC value of our method is 0.9691, which is larger than the AUC of 0.9570 from the state-of-the-art L1E. Furthermore, in the 10-fold cross validation, the AUC value of our method reaches 0.9745, which is also larger than the best result of L1E whose AUC value is 0.9599. Our method also obtain the best prediction performance in the de novo drug validation, its AUC value is 0.9292, which is also larger than the the best result of other methods (WAE (weighted average ensemble method) : 0.9073). In addition, the comparison of the average running time further improves that our method has the higher running efficiency than other competing methods. Finally, the verification results of case studies also prove the prediction ability of our method in practical applications and show that DDI-IS-SL is an effective computational method to predict new DDIs.

2 MATERIALS

Recently, many databases related with drugs have been built based on the experiment validation results of references [40], [41], [42], [43], [44], which include Pubchem, Drugbank, KEGG, SIDER, OFFSIDES and TWOSIDES. Pubchem is the drug chemical substructure information database. Drugbank and KEGG include the drug biological information. SIDER and OFFSIDES provide the drug phenotype information. In this study, we use chemical information, phenotypic data and biological data of drugs to predict DDIs.

The used chemical substructure information of drugs is PubChem substructure fingerprints with 2D binary fingerprints (0 and 1) with the dimension of 881, which are downloaded from the PubChem Compound database [41]. The biological features of drugs contain drug-target interactions, drug enzymes, drug transports and drug pathways. DrugBank is a comprehensive online database containing extensive biochemical, pharmacological information and even quantitative structure activity relationships (QSAR) information about drugs [42], [43], [45]. We extract drug-target interactions, drug enzymes and drug transports from DrugBank. DrugBank was also used to predict drug-drug interaction types [46]. In addition, we also download the drug pathways from KEGG which is a knowledge base for systematic analysis of gene functions [44]. The phenotypic data of drugs are also important features to predict DDIs. In this study, drug-indications, drug-side effects and drug-off side effects construct the phenotypic data of drugs. We extract drug-indications and drug-side effects from SIDER that consists of 1,430 drugs and 5,580 side effect terms. Furthermore, OFF-SIDES is the original data source to provide the drug-off side effects [40].

TABLE 1
The description of datasets

Data type	Data	Database	dimensionality
chemical	Chemical substructures	PubChem	881
Biological	Drug-targets	DrugBank	780
	Drug transporters	DrugBank	18
	Drug enzymes	DrugBank	129
	Drug pathways	KEGG	253
Phenotypic	Drug indications	SIDER	4,897
	Drug side effects	SIDER	4,897
	Drug off side effects	OFFSIDES	9,496
Interaction	Drug-drug interactions	TWOSIDES	DDIs:48,584

Known DDIs are downloaded from TWOSIDES [40]. After projecting drugs of TWOSIDES to PubChem, DrugBank, KEGG, SIDER and OFFSIDER, we obtain a benchmark dataset of known DDIs. It includes 548 drugs and 48,584 DDIs among them. The basic description about data type, data source and dimension of these datasets are demonstrated in Table 1. Furthermore, these datasets also can be downloaded from previous literature [47].

3 METHODS

3.1 Similarity with chemical, biological and phenotypic data

In order to effectively calculate the similarity of feature similarity, a binary vector of any drug is constructed with the obtained chemical, biological and phenotypic data about drugs. Specifically, the values of 1 and 0 in these vector elements indicate the presence or absence of corresponding chemical substructures, targets, transporters, enzymes, pathways, indications, side effects and off side effects features. The feature dimensionality of drugs is 21,351.

After integrating the features of drugs, we calculate drug feature similarity of drugs d_i and d_j as follows:

$$Sim_d(d_i, d_j) = \frac{\sum_{l=1}^{M} F_i(l)F_j(l)}{\sqrt{\sum_{l=1}^{M} F_i^2(l)}\sqrt{\sum_{l=1}^{M} F_j^2(l)}} \quad (1)$$

where $F_i(l)$ and $F_j(l)$ are the l-th elements of feature vectors of drugs d_i and d_j, respectively, and M is equal to 21,351. From Eq(1), the feature similarity of drugs ranges from 0 to 1.

3.2 Gaussian interaction profile kernel similarity

Based on the assumption that drugs exhibiting a similar pattern of interactions and non-interactions with the other drugs in a drug-drug interaction network are similar, we also further compute the GIP kernel similarity of drugs by known DDIs. It also widely used in the other prediction issues, such as drug-target interactions [31], circRNA-disease associations [48], microbe-disease associations [49], lncRNA-disease associations [50], [51] and so on.

Let $D = \{d_1, d_2, ..., d_N\}$ be the set of N drugs. Then the adjacency matrix $Y \in N * N$ represents a known drug-drug interaction network. If drug d_i interacts with drug d_j, the value of y_{ij} is 1 and otherwise is 0. Then we can calculate the Gaussian Interaction Profile (GIP) similarity of drug pairs based on the adjacency matrix Y. Specifically, the GIP kernel similarity of drugs d_i and d_j can be computed as follows:

$$K_{GIP,d}(d_i, d_j) = exp(-\gamma_d||yd_i - yd_j||^2), \tag{2}$$

where the binary vector $yd_i = \{y_{i1}, y_{i2}, ..., y_{iN}\}$ denotes the interaction profile of drug d_i, and γ_d is used to control the kernel bandwidth and is calculated by the interactions profiles of all drugs. It is calculated as follows:

$$\gamma_d = \gamma_d^{'}/(\frac{1}{N}\sum_{i=1}^{N}||yd_i||^2), \tag{3}$$

where $\gamma_d^{'}$ is set to 1 according to previous studies [38], [52], [53].

3.3 RLS classifier and integrating similarities

Before computing the final similarity of drugs, we integrate the feature similarity and GIP similarity of drugs to obtain the final similarity S_d of drugs as follows:

$$S_d = \frac{Sim_d + K_{GIP,d}}{2}, \tag{4}$$

where Sim_d and $K_{GIP,d}$ are the feature similarity of drugs and GIP similarity of drugs, respectively.

The kernel RLS classifier is a typical semi-supervised learning model and has been widely used in other relative prediction issues, such as drug-target interactions and so on. In this study, we also adopt RLS to predict DDIs [38]. Accordingly to the interaction possibility scores of drug-drug pairs are calculated as follows:

$$\hat{Y}_p^T = S_d(S_d + \sigma I)^{-1}Y^T \tag{5}$$

$$\hat{Y} = \frac{\hat{Y}_p + \hat{Y}_p^T}{2} \tag{6}$$

where σ is the regularization parameter, Y is the symmetrical adjacency matrix of known DDIs, S_d is the final similarity matrix of drugs and I is an identity matrix with same size of matrix S_d. $\hat{Y}$ represents the prediction result matrix which is the mean of matrix $\hat{Y}_p$ and its transpose $\hat{Y}_p^T$. We also set the value of σ to 1 according to previous studies [31], [38].

3.4 node-based drug networks diffusion for new drugs

We all know that the prediction ability for new drugs is also important part to evaluate the prediction performance of computational methods. However, some current models are useless to predict DDIs for the new drugs which have no known interactions with other drugs. In this study, inspired by the successful applications of CSN [54] and SDTNBI [55] methods, we also add a preprocess to compute the initial interaction profiles for new drugs. Based on the CSNs of disease phenotypic relationships, Chen et al. proposed a model to predict disease-gene associations, in which disease nodes are connected via one or more shared phenotype nodes. SDTNBI was a substructure-drug-target network-based inference method to prioritize potential targets for old drugs, failed drugs and new chemical entities. It incorporated substructure-drug and drug-target network to bridge the gap between new chemical entities and known DTI network. First, we construct a drug network by integrating drug-target interactions, drug-indications and known DDIs. Then we adopt node-based drug network diffusion model to calculate the relational score for new drugs. When the number of drugs is very large, the node-based drug network is much sparser than the similarity network which is constructed from drug-target interactions, drug-indications and known DDIs. The reason is that although a drug has a small number of targets and indications, it may share at least one target or indication with many other drugs. Moreover, edges of a node-based drug network are based on the observational facts, and thus has less noise than edges in a similarity network. Especially drugs with a small number of targets and indications may share with many other drugs. In addition, in the similarity-based drug network constructed with these biological networks, a seed node can reach each neighbor to be low in the diffusion process, the reason is that they have a larger numbers of neighbors than in the node-based drug network. It also results that the similarity-based drug network is sensitive to noises [54].

Firstly, we construct an adjacency matrix A as follows:

$$A = \begin{bmatrix} Y & M_{dt} & M_{di} \\ M_{dt}^T & 0 & 0 \\ M_{di}^T & 0 & 0 \end{bmatrix} \tag{7}$$

where Y is the adjacency matrix of known DDIs, which includes the new drugs with their interaction profiles of zero-valued vectors. In addition, adjacency matrices M_{dt} and M_{di} represent the known drug-target interactions and known drug indication interactions, respectively. Therefore, A is a systematic matrix with the size of $(N + N_t + N_i) * (N+N_t+N_i)$, where N, N_t and N_i are the number of drugs, targets and indications, respectively.

The node-based drug network diffusion is a two-round resource transfer process. Specifically, in the first transfer process, for a drug d_{new}, the initial resource transferred from its linked target node t_i to drug d_j is calculated as follows:

$$R_{st}(t_i, d_j) = \frac{A(t_i, d_j)}{\sum_{l=1}^{N} A(t_i, l)} * A(t_i, d_{new}), \tag{8}$$

Then drug d_j obtains the allocated resource by adding the contributions from all target nodes and indication nodes associated to it as follows:

$$R_{st}(d_j) = \sum_{l=N+1}^{N+N_t+N_i} R_{st}(l, d_j), \tag{9}$$

In the second transfer process, the resource of drugs obtained in the first round is allocated to drugs by transfer-

ring weights from drugs to drugs. Specifically, the resource transfer from drug d_j to drug d_i is calculated as follows:

$$R_{nd}(d_j, d_i) = \frac{A(d_j, d_i)}{\sum_{l=1}^{N} A(d_i, l)} * R_{st}(d_j), \tag{10}$$

Then the final resource allocated from drug d_{new} to drug d_i can be calculated as follows:

$$R(d_{new}, d_i) = \sum_{l=1}^{N} R_{nd}(l, d_i), \tag{11}$$

where $R_(d_{new}, 1:N)$ represents the values of interaction score vector between drug d_{new} and other drugs. It is used as the the initial scores for new drug d_{new}. We adopt the proportion method to expend these values to higher possibility values by considering that the computed scores via the node-based drug network diffusion method are too small compared to 1 (the value of other known DDIs). The expended process is defined as follows:

$$R(d_{new}, 1:N) = R(d_{new}, 1:N) * (\frac{\alpha}{max(R(d_{new}, 1:N))}), \tag{12}$$

where α is used to control the possibility score that the max value of $R(i, 1:N)$ need to be expanded. In addition, other values are also expanded according to the ratio. In this study, the value of α is set by the de novo drug validation. Fig.1 demonstrates the flow chart of our method.

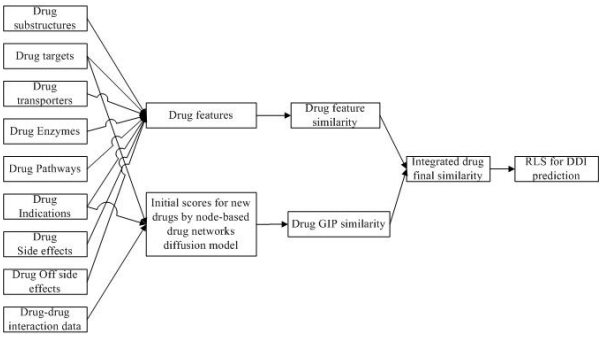

Fig. 1. The flow chart of DDI-IS-SL.

3.5 Algorithm description

Algorithm 1 describes our proposed DDI-IS-SL for predicting DDIs. First, DDI-IS-SL finds the new drug set which has no known DDIs. Second, based on the drug-target interactions, drug-indication interactions and known DDIs, DDI-IS-SL computes the initial interaction scores between these new drugs and other drugs which have known DDIs by the node-based drug network diffusion method. Third, DDI-IS-SL calculates the feature similarity Sim_d of drugs by the cosine measure. The GIP similarity $K_{GIP,d}$ is also computed by the known DDIs. Then DDI-IS-SL further integrates the final similarity of drugs with the mean of Sim_d and $K_{GIP,d}$. Finally, a semi-supervised learning method (RLS) is used to predict DDIs based on the final similarity of drugs and known DDIs. By considering the DDI adjacency matrix is symmetrical, we obtain the final prediction result matrix $\hat{Y}$ via taking the mean of $\hat{Y}_p$ and its transpose $\hat{Y}_p^T$.

Algorithm 1 DDI-IS-SL

Input: Drug set D, drug feature matrices F, M_{dt} and M_{di} , DDIs adjacency matrix Y and parameter α
Output: predicted interaction matrix $\hat{Y}$
DDI-IS-SL(DS,Y,α)

$\sigma = 1$;
Find drug set D_{New} (0 vector) in which each drug has no any DDI interacted with other drugs;
$Y_{old} = Y$; Y_{old} is not updated in the iteration process for new drugs;
for i in D_{New}
  Construct adjacency matrix A based on matrix Y_{old}, M_{dt} and M_{di};
  Construct matrix R by the node-based drug network diffusion method;
  $R(i,:) = R(i,:) * (\frac{\alpha}{max(R(i,:))})$;
  $Y(i,:) = R(i,:)$;
  $Y(:,i) = R(i,:)^T$;
end
Calculate feature similarity Sim_d by drug feature matrix F;
Calculate GIP similarity $K_{GIP,d}$ of drugs;
Calculate the final similarity of drugs by $S_d = \frac{Sim_d + K_{GIP,d}}{2}$;
$\hat{Y}_p^T = S_d(S_d + \sigma I)^{-1} Y^T$;
$\hat{Y} = \frac{\hat{Y}_p + \hat{Y}_p^T}{2}$;
Return $\hat{Y}$;

4 RESULTS AND DISCUSSION

4.1 Benchmark evaluation and evaluation indices

In this study, we conduct the 5-fold cross validation, 10-fold cross validation and de novo drug validation to systematically assess the prediction performance of our method. The AUC is used as the metric. In addition, we also compare our method with other competing DDI prediction methods. In the 5-fold cross validation, the known DDIs are divided into 5 groups and then take turns to use one group as testing samples and the rest as the training samples. Similarly, in the 10-fold cross validation, we also divided the known DDIs into 10 groups and then take turns to use one group as testing samples and the rest as the training samples. The de novo drug validation is used to assess the prediction ability of computational methods for new drugs. One drug is chosen as the test set and the other drugs as the training set in each time of the de novo drug validation, and we conduct the de novo drug validation for all drugs. The AUC value was widely used as the metric to assess the prediction performance of methods. The AUC value of 1 represents the perfect prediction performance of method. The AUC value of less than 0.5 represents the inability of the prediction. Furthermore, we also compare the average computation times of these methods with the 5-fold cross validation.

4.2 Comparison with previous methods

In this study, we compare DDI-IS-SL with other competing methods, namely WAE, L1E, L2E [47] and LP (label propagation method) [33]. WAE, L1E and L2E were three ensem-

TABLE 2
The prediction performances of different methods in 5-fold cross validation, the best result is in the bold face.

Method	Feature	AUC
WAE	Chemical data, biological data, phenotypic data	0.9502
L1E	Chemical data, biological data, phenotypic data	0.9570
L2E	Chemical data, biological data, phenotypic data	0.9561
LP	Drug-sub	0.9356
	Drug-Label	0.9364
	Drug-Off Label	0.9374
DDI-IS-SL	Chemical data, biological data, phenotypic data	**0.9691**

TABLE 3
The prediction performances of different methods in 10-fold cross validation, the best result is in the bold face.

Method	Feature	AUC
WAE	Chemical data, biological data, phenotypic data	0.9530
L1E	Chemical data, biological data, phenotypic data	0.9599
L2E	Chemical data, biological data, phenotypic data	0.9594
LP	Drug-sub	0.9359
	Drug-Label	0.9368
	Drug-Off Label	0.9378
DDI-IS-SL	Chemical data, biological data, phenotypic data	**0.9745**

ble methods that integrate with the neighbor recommend method, random walk method and matrix perturbation method [56]. However, LP was a network-based method to predict DDIs, which only used the drug substructures of chemical data, and drug side effects and drug off side effects of phenotypic data. WAE was a weighted average ensemble model by applying the genetic algorithm (GA) to determine optimal weights, which integrates with the neighbor recommend method, random walk method and matrix perturbation method. L1E and L2E were the classifier ensemble models, which adopt the classifier ensemble rule by a logistic regression classifier with L1 regularization and L2 regularization, respectively.

4.2.1 5-fold cross validation

We obtain the the prediction performances of different methods in the 5-fold cross validation by 10 repeats. Table 2 shows the experiment results of different methods. We can see from Table 2 that our method is superior to other competing methods in terms of AUC values (DDI-IS-SL: 0.9691, WAE: 0.9502, L1E: 0.9570, L2E: 0.9561, LP (max:Drug-Off Label): 0.9374).

4.2.2 10-fold cross validation

We also conduct 10 repeats of the 10-fold cross validation to obtain the prediction performances of our method and other competing methods. Table 3 shows that our method also outperforms other competing methods in terms of AUC. Specifically, the AUC value of our method reaches 0.9745, which is larger than other competing methods (WAE: 0.9530, L1E: 0.9599, L2E: 0.9594 and LP (max:Drug-Off Label): 0.9378).

4.2.3 De novo drug validation

Furthermore, the prediction ability for new drugs of different methods is also compared by the de novo validation. Because other methods can't predict potential DDIs for new drugs, we perform the de novo drug validation experiments on WAE method, Label propagation method and our method. WAE integrates the neighbor recommend method and random walk method. It uses all chemical data, biological data and phenotypic data to compute the similarities, respectively. LP is a network-based method, we also respectively conduct the de novo drug validation on it by three types of features of drugs (Drug-sub, Drug-Label and Drug-Off Label). Fig. 2 shows the experiment results of DDI-IS-SL, WAE, LP (LP-sub: drug-sub, LP-lab: drug-label, LP-off: drug-off label). We also can see from Fig. 2 that the AUC values of our method achieves 0.9292, which is also better than other methods (WAE: 0.9073, LP (max:Drug-Off Label): 0.8997).

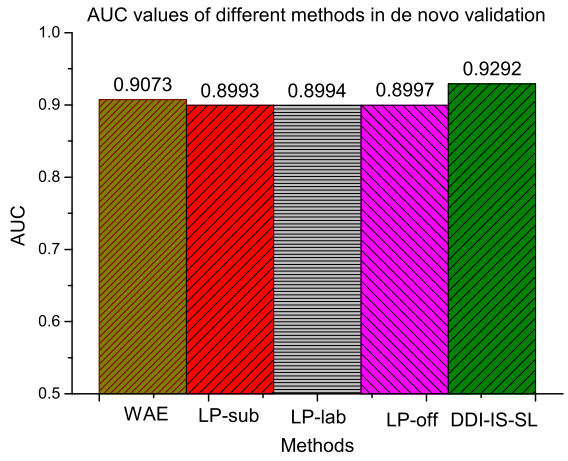

Fig. 2. The prediction performances of different methods in de novo validation.

4.2.4 Comparison of the average computation time

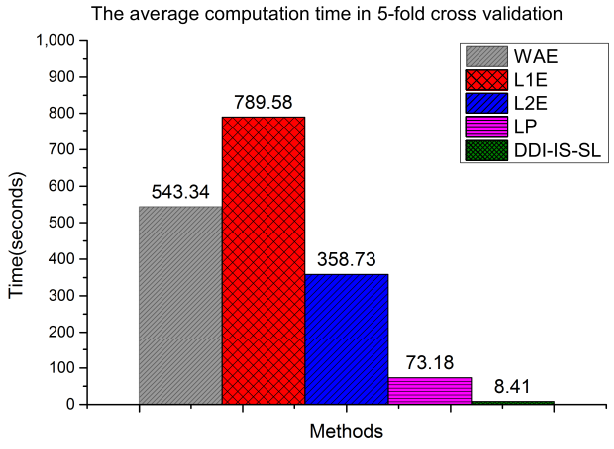

Fig. 3. The average computation time of different methods in 5-fold cross validation.

The computation time is also an important metric to assess the performance of computational methods. Therefore, we compare the average running time of our method with these of other methods. The average computation time of each method was obtained from 10 repeats of 5-fold cross

validation on a Windows 7 64bit with an Intel Core i3 3.4GHZ, dual cores, 12GB of RAM. Fig.3 shows the average computation time of all computational methods. We can see from Fig.3 that the computation time of our method is 8.41 seconds, which is much less than those of other methods. In addition, the average computation time of LP is longer than that of our method and less than those of WAE, L1E and L2E. The average computation times of WAE, L1E and L2E are far longer than those of LP and our DDI-IS-SL. The reason is that WAE, L1E and L2E are the integrated methods. Especially, the average computation time of L1E reaches 789.58 seconds.

Generally, based on experiment results of the 5-fold cross validation, 10-fold cross validation, de novo drug validation and computation time comparison, our method outperforms other competing methods. In addition, we can also see from Table 2 and Table 3 that the advantage of DDI-IS-SL over other methods in the 10-fold cross validation is more obvious than that in the 5-fold cross validation. In the 5-fold cross validation, the AUC value of DDI-IS-SL is 0.9691, while the best result of other methods is 0.9570. However, in the 10-fold cross validation, the AUC value of DDI-IS-SL is 0.9745, while the best result of other methods is 0.9599. Their improvements are 0.0121 and 0.0146, respectively. It also demonstrates that the DDI-IS-SL is more effective to predict DDIs in real applications, when there are many known DDIs. Moveover, by comparing the de novo drug validation and the average computation time, the advantages of our method are further demonstrated.

4.3 Parameter analysis for α

For new drugs, we take the node-based drug network diffusion model to compute the initial interaction profiles and also use the parameter α to control the expanded possibility scores of the maximum value. Therefore, we set the value of parameter α based on experiment results of the de novo drug validation.

TABLE 4
The AUC of DDI-IS-SL under different settings of α, the default result is in the bold face.

α	α = 0	α =0.1	α =0.2	α =0.3	α =0.4
AUC	0.7783	0.9005	0.9221	0.9275	0.9290
α =0.5	α =0.6	α =0.7	α =0.8	α =0.9	α =1.0
0.9292	0.9289	0.9285	0.9280	0.9275	0.9271

Table 4 demonstrates that the sensitivity of DDI-IS-SL with different values of parameter α. When the value of α is set to 0, the prediction performance of our method is worst. The reason is that $\alpha = 0$ is equal to directly predict DDIs without adding an initial process in the de novo validation, which indicates that the node-based networks diffusion method is very effective to predict DDIs for new drugs. In addition, we can also see from Table 4 that the values of α are from 0.3 to 1.0 has little effect on the prediction performance of our method. The prediction performance of DDI-IS-SL has an increasing trend when the values of α from 0 to 0.3. When the value of α is set to be 0.5, our method achieves the best prediction performance (AUC: 0.9292). Table 4 indicates that the expanding method can effectively improve the prediction performance of our method in the de novo drug validation.

4.4 Iteration times of node-based drug network diffusion model

In the node-based drug network diffusion model, the drug-substructure network and drug-target interaction network are used in the first transfer process, and the drug-drug interaction network is used in the second transfer process. Therefore, in order to full use the drug-substructure, drug-target interaction network and drug-drug interaction network information, we set that the node-based drug networks diffusion model includes two resource transfer processes. However, we can also compute the initial interaction scores for new drugs when using the first transfer process, and the AUC value is 0.7954. In addition, we can also compute the initial interaction scores for new drugs when running the node-based drug networks diffusion model multiple times. Table shows the experiment results of DDI-IS-SL when conducting the node-based drug network diffusion model K times. We can see from Table 7 that DDI-IS-SL obtains the best prediction performance when $K = 1$. Therefore, we running the node-based drug network diffusion one time in DDI-IS-SL.

4.5 Feature combination analysis

In this section, we analyze the prediction performances of different drug features in our method. We take the 10-fold cross validation and the de novo drug validation to choose final drug features to predict DDIs.

Table 6 demonstrates that the prediction performances of DDI-IS-SL with different drug feature combinations in terms of sensitivity. We can see from Table 6 that DDI-IS-SL can obtain the reliable prediction performance when any drug feature is combined with GIP. However, the AUC value of DDI-IS-SL is 0.9617 when only using GIP. The prediction performance is the best (AUC:0.9745) when all drug features and GIP are combined. In addition, our method also can obtain the best prediction performance when drug off side effects and GIP are combined. We think that the combination of drug features and GIP can improve the prediction performance of DDI-IS-SL. Therefore, we choose the combination of all drug features and GIP to predict DDIs.

By considering biological networks of CSN and SDTNBI, we choose the final drug feature networks to predict DDIs for new drugs by the de novo drug validation. The final drug feature networks are selected from the drug chemical information (drug-substructures), drug biological information (drug-target interactions) and phenotypic information (drug indications). Table 7 demonstrates that DDI-IS-SL can obtain the best prediction performance (AUC:0.9292) when using drug-targets and drug indications. Based on the experiment results of de novo drug validation, we think that the combination of drug-targets and drug indications can improve our prediction performance of DDI-IS-SL. Therefore, we choose drug-substructures and drug indications to compute initial interaction scores for new drugs on node-based drug network diffusion process.

TABLE 5
The prediction performances of DDI-IS-SL when running node-based drug networks diffusion model K times on the de novo drug validation.

K	K=1	K=2	K=3	K=4
AUC	**0.9292**	0.9218	0.9164	0.9145

TABLE 6
The prediction performances of DDI-IS-SL with different drug feature combinations on the 10-fold cross validation.

No	Feature combination	AUC
1	Drug-substructures and GIP	0.9743
2	Drug-targets and GIP	0.9738
3	Drug transporters and GIP	0.9739
4	Drug enzymes and GIP	0.9735
5	Drug pathways and GIP	0.9736
6	Drug indication and GIP	0.9744
7	Drug side effects and GIP	0.9743
8	Drug off side effects and GIP	**0.9745**
9	GIP	0.9617
10	All drug features and GIP	**0.9745**

TABLE 7
The prediction performances of DDI-IS-SL with different drug feature combinations on the de novo drug validation.

No	Feature combination	AUC
1	Drug-substructures	0.9163
2	Drug-targets	0.9006
3	Drug indications	0.9287
4	Drug-substructures and Drug-targets	0.9169
5	Drug-substructures and Drug indications	0.9177
6	Drug-targets and Drug indications	**0.9292**
7	Drug-substructures, Drug-targets and Drug indications	0.9182

TABLE 8
Top 20 new DDIs predicted by DDI-IS-SL.

Rank	Drug ID1	Drug ID2	Drug name1	Drug name2	Evidence
1	DB00448	DB01059	Lansoprazole	Norfloxacin	Unknown
2	DB00333	DB00213	Methadone	Pantoprazole	DrugBank
3	DB00991	DB00231	Oxaprozin	Temazepam	Unknown
4	DB00813	DB00535	Fentanyl	Cefdinir	Unknown
5	DB00863	DB00690	Ranitidine	Flurazepam	Unknown
6	DB00470	DB00331	Dronabinol	Metformin	Unknown
7	DB00989	DB01136	Rivastigmine	Carvedilol	DrugBank
8	DB00257	DB00230	Clotrimazole	Pregabalin	DrugBank
9	DB00869	DB00537	Dorzolamide	Ciprofloxacin	Unknown
10	DB00887	DB00207	Bumetanide	Azithromycin	Unknown
11	DB00584	DB00790	Enalapril	Perindopril	DrugBank
12	DB00264	DB01193	Metoprolol	Acebutolol	DrugBank
13	DB00795	DB00334	Sulfasalazine	Olanzapine	Unknown
14	DB01149	DB00586	Nefazodone	Diclofenac	DrugBank
15	DB00327	DB00252	Hydromorphone	Phenytoin	DrugBank
16	DB01190	DB01132	Clindamycin	Pioglitazone	Unknown
17	DB00328	DB00218	Indomethacin	Moxifloxacin	Unknown
18	DB00549	DB01129	Zafirlukast	Rabeprazole	DrugBank
19	DB00158	DB01026	Folic Acid	Ketoconazole	Unknown
20	DB01203	DB00335	Nadolol	Atenolol	DrugBank

4.6 Case studies

In order to further validate the prediction performance of our method in the practical applications, we verify its top 20 newly predicted DDIs. The benchmark dataset composes of 548 drugs and 48,584 known DDIs and it is downloaded from TWOSIDES. The predicted DDIs are verified in DrugBank. Table 8 shows that 9 of top 20 DDIs are confirmed. For example, Methadone is a synthetic opioid that is used as the hydrochloride and as part of the treatment of dependence on opioid drugs. When it is combined with Pantoprazole, its metabolism can be decreased [42]. Furthermore, Carvedilol is a non-selective beta blocker indicated in the treatment of mild to moderate congestive heart failure (CHF), it also increases the bradycardic activities of Rivastigmine [45]. Clotrimazole is an imidazole derivative with a broad spectrum of antimycotic activity while Pregabalin is an anticonvulsant drug used for neuropathic pain, epilepsy and generalized anxiety disorder, and their combination can increased the risk or severity of adverse effects [57]. Similarly, the metabolism is decreased when Diclofenac is combined with Nefazodone, Hydromorphone is combined with Phenytoin and Zafirlukast is combined with Rabepra-

zole [43]. Meanwhile, other predicted DDIs have not been validated in DrugBank, which is worth to study in the future by biomedical experiments.

In addition, we also verify the top 20 predicted DDIs of Imipramine in the de novo validation. They are confirmed in TWOSIDES and DrugBank. Imipramine is a dibenzazepine-derivative TCA (the prototypical tricyclic antidepressant) and is used to treat depression and nocturnal enuresis in children [58]. Table 9 also demonstrates that top 20 predicted DDIs are validated in TWOSIDES and 18 out of top 20 predicted DDIs are verified in DrugBank. When Imipramine is combined with one of Lorazepam, Citalopram, Clonazepam, Furosemide, Paroxetine, Sertraline, Alprazolam, Diazepam and Salbutamol, the risk or severity of adverse effects is increased [42], [58]. Furthermore, when it is combined with one of Acetaminophen, Omeprazole, Ranitidine, Simvastatin and Pantoprazole, the metabolism of Imipramine is decreased [59]. Zolpidem is a prescription short-acting nonbenzodiazepine hypnotic that potentiates gamma-aminobutyric acid (GABA) and is used for the short-term treatment of insomnia, its central nervous system depressant activities can be increased with Imipramine [60]. Levothyroxine is the major hormone derived from the thyroid gland, it can increase the arrhythmogenic activities of Imipramine [45], [61]. Metoprolol is widely used for acute myocardial infarction (MI), heart failure, angina pectoris and mild-to-moderate hypertension, its serum concentration can be increased when Metoprolol is combined with Imipramine [62]. Similarly, its serum concentration can also be increased when Lisinopril is combined with Imipramine [63].

The top 20 predicted DDIs of Trandolapril resulted from the de novo validation are also validated in TWOSIDES and DrugBank. Trandolapril is a non-sulhydryl prodrug that may be used to treat mild-to-moderate hypertension, and slow the rate of progression of renal disease in hypertensive individuals with diabetes mellitus and microalbuminuria or overt nephropathy [64]. We can see from Table 10 that 19 out of top 20 predicted DDIs are verified by TWOSIDES or DrugBank. For example, the hyperkalemic activities of Trandolapril can be increased by combining Ardeparin which is derived via the peroxide degradation of heparin extracted from porcine intestinal mucosa and is used for the prevention of postoperative venous thrombosis [42], [65]. The serum concentration of Simvastatin or Digoxin can be increased when they combine with Trandolapril [66]. In addition, the antihypertensive activities of Trandolapril may also be decreased by Acetylsalicylic acid which is the prototypical analgesic used in the treatment of mild to moderate pain [67]. The risk or severity of adverse effects can be increased when Trandolapril is combined with one of Furosemide, Metoprolol, Lisinopril and Hydrochlorothiazide [45].

5 Conclusion

Multi-drug therapies have widely been used to treat diseases, especially complex diseases such as cancer to improve the treatment effect and reduce the burden of patients. However, the adverse effects resulted from multi-drug therapies have also been observed, which may caused some serious complications and even the patient death. Therefore, identifying drug-drug interactions is helpful in contributing to improved treatment of diseases and reducing the difficulty of drug developments. Especially, it is very necessary to develop new computational methods for identifying DDIs.

In this study, we propose a new computational method (DDI-IS-SL) to infer DDIs. DDI-IS-SL integrates the drug chemical, drug biological and drug phenotypic data. The used chemical substructure information of drugs is PubChem substructure which is the 2D binary fingerprints (0 and 1). The biological features of drugs contain drug-target interactions, drug enzymes, drug transports and drug pathways. The phenotypic data of drugs include drug-indications, drug-side effects and drug-off side effects. For each drug, a high-dimensional binary feature vector is constructed with these data. Then we calculate the feature similarity of drugs with the cosine measure. We also compute the GIP similarity of drugs by known DDIs. The final similarity of drugs is calculated as the mean of drug feature similarity and drug GIP similarity. Then we use a semi-supervised learning model (RLS) to compute the probability scores of drug pairs. In the 5-fold cross validation and 10-fold cross validation, DDI-IS-SL achieves the better prediction performance than other competing methods. Furthermore, for new drugs, we also calculate the relational initial interaction scores by using the node-based drug network diffusion method. Our method also achieves the better prediction performance in de novo validation than competing methods.

Although the DDI-IS-SL is an effective approach to predict the potential DDIs, there are still some areas for the improvement. For example, we can also consider other more sophisticated methods to integrate the chemical, biological and phenotypic data of drugs. In addition, other prediction models such as deep learning method and matrix approximation method [68], [69], [70] also should be tried to identify DDIs in the future.

Acknowledgment

The authors are very grateful to the anonymous reviewers for their constructive comments which have helped significantly in revising this work. The authors would like to express their gratitude for the support from the National Natural Science Foundation of China(No.61772552, No.61420106009, No.61832019 and No.61962050), 111 Project (No. B18059) and Hunan Provinvial Science and Technology Program (No. 2018WK4001).

References

[1] D. Quinn and R. Day, "Drug interactions of clinical importance," *Drug safety*, vol. 12, no. 6, pp. 393–452, 1995.

[2] T. Prueksaritanont, X. Chu, C. Gibson, D. Cui, K. L. Yee, J. Ballard, T. Cabalu, and J. Hochman, "Drug–drug interaction studies: Regulatory guidance and an industry perspective," *The AAPS journal*, vol. 15, no. 3, pp. 629–645, 2013.

[3] H. Kusuhara, "How far should we go? perspective of drug-drug interaction studies in drug development," *Drug metabolism and pharmacokinetics*, vol. 29, no. 3, pp. 227–228, 2014.

[4] N. R. Crowther, A. M. Holbrook, R. Kenwright, and M. Kenwright, "Drug interactions among commonly used medications. chart simplifies data from critical literature review." *Canadian Family Physician*, vol. 43, p. 1972, 1997.

TABLE 9
The validation result of top 20 new DDIs of drug Imipramine predicted by DDI-IS-SL method in the de novo validation.

Rank	Drug ID1	Drug ID2	Drug name1	Drug name2	Evidence
1	DB00458	DB00338	Imipramine	Omeprazole	TWOSIDES,DrugBank
2		DB00186		Lorazepam	TWOSIDES,DrugBank
3		DB00425		Zolpidem	TWOSIDES,DrugBank
4		DB00316		Acetaminophen	TWOSIDES,DrugBank
5		DB00451		Levothyroxine	TWOSIDES,DrugBank
6		DB00215		Citalopram	TWOSIDES,DrugBank
7		DB00863		Ranitidine	TWOSIDES,DrugBank
8		DB00641		Simvastatin	TWOSIDES,DrugBank
9		DB00448		Lansoprazole	TWOSIDES
10		DB01068		Clonazepam	TWOSIDES,DrugBank
11		DB00695		Furosemide	TWOSIDES,DrugBank
12		DB00715		Paroxetine	TWOSIDES,DrugBank
13		DB01104		Sertraline	TWOSIDES,DrugBank
14		DB00404		Alprazolam	TWOSIDES,DrugBank
15		DB00829		Diazepam	TWOSIDES,DrugBank
16		DB00264		Metoprolol	TWOSIDES,DrugBank
17		DB01050		Ibuprofen	TWOSIDES
18		DB00722		Lisinopril	TWOSIDES,DrugBank
19		DB00213		Pantoprazole	TWOSIDES,DrugBank
20		DB01001		Salbutamol	TWOSIDES,DrugBank

TABLE 10
The validation result of top 20 new DDIs of drug Trandolapril predicted by DDI-IS-SL method in the de novo validation.

Rank	Drug ID1	Drug ID2	Drug name1	Drug name2	Evidence
1	DB00519	DB00338	Trandolapril	Omeprazole	TWOSIDES
2		DB00316		Acetaminophen	TWOSIDES
3		DB00451		Levothyroxine	TWOSIDES
4		DB00641		Simvastatin	TWOSIDES,DrugBank
5		DB00695		Furosemide	TWOSIDES,DrugBank
6		DB00945		Acetylsalicylic acid	TWOSIDES,DrugBank
7		DB00863		Ranitidine	TWOSIDES
8		DB00448		Lansoprazole	TWOSIDES
9		DB00758		Clopidogrel	TWOSIDES
10		DB00264		Metoprolol	TWOSIDES,DrugBank
11		DB00722		Lisinopril	DrugBank
12		DB00390		Digoxin	TWOSIDES,DrugBank
13		DB00425		Zolpidem	TWOSIDES
14		DB00215		Citalopram	TWOSIDES
15		DB00213		Pantoprazole	Unknown
16		DB00186		Lorazepam	TWOSIDES
17		DB00927		Famotidine	TWOSIDES
18		DB00407		Ardeparin	TWOSIDES,DrugBank
19		DB00635		Prednisone	TWOSIDES
20		DB00999		Hydrochlorothiazide	TWOSIDES

[5] R. Nahta, M.-C. Hung, and F. J. Esteva, "The her-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells," *Cancer research*, vol. 64, no. 7, pp. 2343–2346, 2004.

[6] T.-C. Chou, "Drug combination studies and their synergy quantification using the chou-talalay method," *Cancer research*, vol. 70, no. 2, pp. 440–446, 2010.

[7] K. Venkatakrishnan, L. L. von Moltke, R. Obach, and D. J. Greenblatt, "Drug metabolism and drug interactions: application and clinical value of in vitro models," *Current drug metabolism*, vol. 4, no. 5, pp. 423–459, 2003.

[8] P. J. Neuvonen, M. Niemi, and J. T. Backman, "Drug interactions with lipid-lowering drugs: mechanisms and clinical relevance," *Clinical Pharmacology & Therapeutics*, vol. 80, no. 6, pp. 565–581, 2006.

[9] Y. Böttiger, K. Laine, M. L. Andersson, T. Korhonen, B. Molin, M.-L. Ovesjö, T. Tirkkonen, A. Rane, L. L. Gustafsson, and B. Eiermann, "Sfinxła drug-drug interaction database designed for clinical decision support systems," *European journal of clinical pharmacology*, vol. 65, no. 6, pp. 627–633, 2009.

[10] M. P. Pai, D. M. Graci, and G. W. Amsden, "Macrolide drug interactions: an update," *Annals of Pharmacotherapy*, vol. 34, no. 4, pp. 495–513, 2000.

[11] J. Kuhlmann and W. Mück, "Clinical-pharmacological strategies to assess drug interaction potential during drug development," *Drug safety*, vol. 24, no. 10, pp. 715–725, 2001.

[12] S. Preskorn and S. Werder, "Detrimental antidepressant drug–drug interactions: Are they clinically relevant?" *Neuropsychopharmacology*, vol. 31, no. 8, pp. 1605–1612, 2006.

[13] D. Sridhar, S. Fakhraei, and L. Getoor, "A probabilistic approach for collective similarity-based drug–drug interaction prediction," *Bioinformatics*, vol. 32, no. 20, pp. 3175–3182, 2016.

[14] S. Ekins and S. A. Wrighton, "Application of in silico approaches to predicting drug–drug interactions," *Journal of pharmacological and toxicological methods*, vol. 45, no. 1, pp. 65–69, 2001.

[15] G. Jin, H. Zhao, X. Zhou, and S. T. Wong, "An enhanced petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data," *Bioinformatics*, vol. 27, no. 13, pp. i310–i316, 2011.

[16] R. Ferdousi, R. Safdari, and Y. Omidi, "Computational prediction of drug-drug interactions based on drugs functional similarities," *Journal of Biomedical Informatics*, vol. 70, pp. 54–64, 2017.

[17] R. Safdari, R. Ferdousi, K. Aziziheris, S. R. Niakan-Kalhori, and Y. Omidi, "Computerized techniques pave the way for drug-drug interaction prediction and interpretation," *BioImpacts: BI*, vol. 6, no. 2, p. 71, 2016.

[18] N. P. Tatonetti, G. H. Fernald, and R. B. Altman, "A novel signal detection algorithm for identifying hidden drug-drug interactions

in adverse event reports," *Journal of the American Medical Informatics Association*, vol. 19, no. 1, pp. 79–85, 2011.

[19] A. Gottlieb, G. Y. Stein, Y. Oron, E. Ruppin, and R. Sharan, "Indi: a computational framework for inferring drug interactions and their associated recommendations," *Molecular systems biology*, vol. 8, no. 1, p. 592, 2012.

[20] S. Yamazaki, T. R. Johnson, and B. J. Smith, "Prediction of drug-drug interactions with crizotinib as the cyp3a substrate using a physiologically based pharmacokinetic model," *Drug Metabolism and Disposition*, vol. 43, no. 10, pp. 1417–1429, 2015.

[21] L. Tari, S. Anwar, S. Liang, J. Cai, and C. Baral, "Discovering drug–drug interactions: a text-mining and reasoning approach based on properties of drug metabolism," *Bioinformatics*, vol. 26, no. 18, pp. i547–i553, 2010.

[22] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, and C. Friedman, "Drugłdrug interaction through molecular structure similarity analysis," *Journal of the American Medical Informatics Association*, vol. 19, no. 6, pp. 1066–1074, 2012.

[23] S. Vilar, E. Uriarte, L. Santana, T. Lorberbaum, G. Hripcsak, C. Friedman, and N. P. Tatonetti, "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nature protocols*, vol. 9, no. 9, pp. 2147–2163, 2014.

[24] F. Cheng and Z. Zhao, "Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties," *Journal of the American Medical Informatics Association*, vol. 21, no. e2, pp. e278–e286, 2014.

[25] P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, C. Zheng, Z. Guo, X. Chen, W. Zhou *et al.*, "Large-scale exploration and analysis of drug combinations," *Bioinformatics*, vol. 31, no. 12, pp. 2007–2016, 2015.

[26] L. Liu, L. Chen, Y.-H. Zhang, L. Wei, S. Cheng, X. Kong, M. Zheng, T. Huang, and Y.-D. Cai, "Analysis and prediction of drug–drug interaction by minimum redundancy maximum relevance and incremental feature selection," *Journal of Biomolecular Structure and Dynamics*, vol. 35, no. 2, pp. 312–329, 2017.

[27] H. Luo, P. Zhang, H. Huang, J. Huang, E. Kao, L. Shi, L. He, and L. Yang, "Ddi-cpi, a server that predicts drug–drug interactions through implementing the chemical–protein interactome," *Nucleic acids research*, vol. 42, no. W1, pp. W46–W52, 2014.

[28] T. Takeda, M. Hao, T. Cheng, S. H. Bryant, and Y. Wang, "Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge," *Journal of Cheminformatics*, vol. 9, no. 1, p. 16, 2017.

[29] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "Nllss: predicting synergistic drug combinations based on semi-supervised learning," *PLoS computational biology*, vol. 12, no. 7, p. e1004975, 2016.

[30] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, and Y. Pan, "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.

[31] C. Yan, J. Wang, W. Lan, F.-X. Wu, and Y. Pan, "Sdtrls: Predicting drug-target interactions for complex diseases based on chemical substructures," *Complexity*, vol. 2017, 2017.

[32] D.-S. Cao, N. Xiao, Y.-J. Li, W.-B. Zeng, Y.-Z. Liang, A.-P. Lu, Q.-S. Xu, and A. Chen, "Integrating multiple evidence sources to predict adverse drug reactions based on a systems pharmacology model," *CPT: pharmacometrics & systems pharmacology*, vol. 4, no. 9, pp. 498–506, 2015.

[33] P. Zhang, F. Wang, J. Hu, and R. Sorrentino, "Label propagation prediction of drug-drug interactions based on clinical side effects," *Scientific reports*, vol. 5, 2015.

[34] J. Huang, C. Niu, C. D. Green, L. Yang, H. Mei, and J.-D. J. Han, "Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network," *PLoS computational biology*, vol. 9, no. 3, p. e1002998, 2013.

[35] K. Park, D. Kim, S. Ha, and D. Lee, "Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks," *PloS one*, vol. 10, no. 10, p. e0140816, 2015.

[36] B. Jin, H. Yang, C. Xiao, P. Zhang, X. Wei, and F. Wang, "Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction," in *AAAI*, pp. 1367–1373, 2017.

[37] S. Vilar, E. Uriarte, L. Santana, N. P. Tatonetti, and C. Friedman, "Detection of drug-drug interactions by modeling interaction profile fingerprints," *PloS one*, vol. 8, no. 3, p. e58321, 2013.

[38] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug–target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.

[39] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.

[40] N. P. Tatonetti, P. Y. Patrick, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Science translational medicine*, vol. 4, no. 125, pp. 125ra31–125ra31, 2012.

[41] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "Pubchem: a public information system for analyzing bioactivities of small molecules," *Nucleic acids research*, vol. 37, no. suppl_2, pp. W623–W633, 2009.

[42] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "Drugbank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic acids research*, vol. 34, no. suppl_1, pp. D668–D672, 2006.

[43] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu *et al.*, "Drugbank 3.0: a comprehensive resource for omics research on drugs," *Nucleic acids research*, vol. 39, no. suppl_1, pp. D1035–D1041, 2010.

[44] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "Kegg for representation and analysis of molecular networks involving diseases and drugs," *Nucleic acids research*, vol. 38, no. suppl_1, pp. D355–D360, 2009.

[45] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu *et al.*, "Drugbank 4.0: shedding new light on drug metabolism," *Nucleic acids research*, vol. 42, no. D1, pp. D1091–D1097, 2013.

[46] H. Luo, J. Wang, C. Yan, M. Li, W. Fangxiang, and P. Yi, "A novel drug repositioning approach based on collaborative metric learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, to be published. DOI: 10.1109/TCBB.2019.2926453.

[47] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC bioinformatics*, vol. 18, no. 1, p. 18, 2017.

[48] C. Yan, J. Wang, and F.-X. Wu, "Dwnn-rls: regularized least squares method for predicting circrna-disease associations," *BMC bioinformatics*, vol. 19, no. 19, p. 520, 2018.

[49] C. Yan, G. Duan, F. Wu, Y. Pan, and J. Wang, "Brwmda: Predicting microbe-disease associations based on similarities and bi-random walk on disease and microbe networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, to be published. DOI: 10.1109/TCBB.2019.2907626.

[50] W. Lan, M. Li, K. Zhao, J. Liu, F.-X. Wu, Y. Pan, and J. Wang, "Ldap: a web server for lncrna-disease association prediction," *Bioinformatics*, vol. 33, no. 3, pp. 458–460, 2016.

[51] C. Lu, M. Yang, F. Luo, F.-X. Wu, M. Li, Y. Pan, Y. Li, and J. Wang, "Prediction of lncrna-disease associations based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 19, pp. 3357–3364, 2018.

[52] C. Yan, J. Wang, P. Ni, W. Lan, F. X. Wu, and Y. Pan, "Dnrlmf-mda:predicting microrna-disease associations based on similarities of micrornas and diseases," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 1, pp. 233–243, 2019.

[53] W. Lan, J. Wang, M. Li, J. Liu, F.-X. Wu, and Y. Pan, "Predicting microrna-disease associations based on improved microrna and disease similarities," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 6, pp. 1774–1782, 2018.

[54] Y. Chen and R. Xu, "Context-sensitive network-based disease genetics prediction and its implications in drug discovery," *Bioinformatics*, vol. 33, no. 7, pp. 1031–1039, 2017.

[55] Z. Wu, F. Cheng, J. Li, W. Li, G. Liu, and Y. Tang, "Sdtnbi: an integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning," *Briefings in bioinformatics*, vol. 18, no. 2, pp. 333–347, 2016.

[56] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley, "Toward link predictability of complex networks," *Proceedings of the National Academy of Sciences*, vol. 112, no. 8, pp. 2325–2330, 2015.

[57] K. DelPrete, "Topical compositions with long lasting effect," uS Patent App. 11/775,589, 2007.

[58] A. Ledeti, G. Vlase, I. Ledeti, T. Vlase, P. Matusz, C. Dehelean, D. Circioban, L. Stelea, and L. M. Suta, "Thermal stability of desipramine and imipramine," *children*, vol. 1, p. 3, 2016.

[59] M. D. Rawlins, "Variability in response to drugs," *British medical journal*, vol. 4, no. 5936, p. 91, 1974.
[60] E. Krylova, S. Kataev, Y. A. Khomov, and O. Dvorskaya, "Validation of a gas chromatography/mass spectrometry method for quantitative determination of zolpidem in whole blood," *Pharmaceutical Chemistry Journal*, vol. 49, no. 8, pp. 564–569, 2015.
[61] J. J. Dunn, S. T. Columbus, W. E. Aldeen, M. Davis, and K. C. Carroll, "Trichuris vulpis recovered from a patient with chronic diarrhea and five dogs," *Journal of Clinical Microbiology*, vol. 40, no. 7, pp. 2703–2704, 2002.
[62] J. Swen, M. Nijenhuis, A. d. Boer, L. Grandia, A.-H. Maitland-van der Zee, H. Mulder, G. Rongen, R. Schaik, T. Schalekamp, D. Touw *et al.*, "Pharmacogenetics: from bench to bytełan update of guidelines," *Clinical Pharmacology & Therapeutics*, vol. 89, no. 5, pp. 662–673, 2011.
[63] S. E. Melanson, C. P. Stowell, J. G. Flood, E. L. Lewandrowski, R. J. Zak, and K. B. Lewandrowski, "Does blood donor history accurately reflect the use of prescription medications? a comparison of donor history and serum toxicologic analysis," *Transfusion*, vol. 46, no. 8, pp. 1402–1407, 2006.
[64] O. D. Pedersen, H. Bagger, L. Køber, C. Torp-Pedersen *et al.*, "Trandolapril reduces the incidence of atrial fibrillation after acute myocardial infarction in patients with left ventricular dysfunction," *Circulation*, vol. 100, no. 4, pp. 376–380, 1999.
[65] G. L. Mosher, S. G. Machatha, and D. J. Cushing, "Pharmaceutical compositions comprising prasugrel and cyclodextrin derivatives and methods of making and using the same," uS Patent App. 15/181,190, 2016.
[66] V. Athyros, D. Mikhailidis, A. Papageorgiou, V. Bouloukos, A. Pehlivanidis, A. Symeonidis, and M. Elisaf, "Effect of statins and ace inhibitors alone and in combination on clinical outcome in patients with coronary heart disease," *Journal of human hypertension*, vol. 18, no. 11, pp. 781–788, 2004.
[67] B. Takkouche, M. Etminan, F. Caamaño, and P. A. Rochon, "Interaction between aspirin and ace inhibitors," *Drug safety*, vol. 25, no. 5, pp. 373–378, 2002.
[68] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinformatics*, vol. 34, no. 11, pp. 1904–1912, 2018.
[69] C. Yan, G. Duan, F. Wu, Y. Pan, and J. Wang, "Mchmda: Predicting microbe-disease associations based on similarities and low-rank matrix completion," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, to be published. DOI: 10.1109/TCBB.2019.2926716.
[70] M. Yang, H. Luo, Y. Li, and J. Wang, "Drug repositioning based on bounded nuclear norm regularization," *Bioinformatics*, vol. 35, no. 14, pp. i455–i463, 2019.

Cheng Yan received PhD degree in Computer Science and Technology from Central South University in 2018. He is currently an postdoctoral in the School of Computer Science and engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include bioinformatics and machine learning.

Guihua Duan received PhD degree in Computer Science and Technology from Central South University in 2010. She is currently an Associate Professor in the School of Computer Science and engineering, Central South University. Her research interests include network security, bioinformatics and computer network.

Fang-Xiang Wu (M'06-SM'11) received the B.Sc. degree and the M.Sc. degree in applied mathematics, both from Dalian University of Technology, Dalian, China, in 1990 and 1993, respectively, the first Ph.D. degree in control theory and its applications from Northwestern Polytechnical University, Xi'an, China, in 1998, and the second Ph.D. degree in biomedical engineering from University of Saskatchewan (U of S), Saskatoon, Canada, in 2004. During 2004C2005, he worked as a Postdoctoral Fellow in the Laval University Medical Research Center (CHUL), Quebec City, Canada. He is currently a Professor of the Division of Biomedical Engineering and the Department of Mechanical Engineering at the U of S. His current research interests include Artificial Intelligence, Machine/Deep Learning, Computational Biology and Bioinformatics, Medical Image Analytics, Complex Network Analytics. He has published more than 380 technical papers in refereed journals and conference proceedings. Dr. Wu is serving as the editorial board member of four international journals, the guest editor of several international journals, and as the program committee chair or member of many international conferences. He has also reviewed papers for many international journals.

Yi Pan is a Regents' Professor of Computer Science and an Interim Associate Dean and Chair of Biology at Georgia State University, USA. Dr. Pan joined Georgia State University in 2000 and was promoted to full professor in 2004, named a Distinguished University Professor in 2013 and designated a Regents' Professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. He served as the Chair of Computer Science Department from 2005-2013. He is also a visiting Changjiang Chair Professor at Central South University, China. Dr. Pan received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pitts-burgh, USA, in 1991. His profile has been featured as a distin-guished alumnus in both Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. Dr. Pan's re-search interests include parallel and cloud computing, wireless networks, and bioinformatics. Dr. Pan has published more than 330 papers including over 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edit-ed/authored 40 books. His work has been cited more than 6500 times. Dr. Pan has served as an editor-in-chief or editori-al board member for 15 journals including 7 IEEE Transactions. He is the recipient of many awards including IEEE Transac-tions Best Paper Award, 4 other international conference or journal Best Paper Awards, 4 IBM Faculty Awards, 2 JSPS Sen-ior Invitation Fellowships, IEEE BIBE Outstanding Achieve-ment Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.

Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the chair of and a professor in Department of Computer Science, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, Bioinformatics and computer network. He is a senior member of the IEEE.