DPDDI:a Deep Predictor for Drug-Drug Interactions

Yue-Hua Feng
Northwestern Polytechnical University
Shao-Wu Zhang (zhangsw@nwpu.edu.cn)
Northwestern Polytechnical University https://orcid.org/0000-0003-1305-7447
Jian-Yu Shi
Northwestern Ploytechnical University

Methodology article

Keywords: Drug-drug interaction, DDI prediction, graph convolution network (GCN), feature extraction, deep neural network

Posted Date: June 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17012/v2

Version of Record: A version of this preprint was published on September 24th, 2020. See the published version at https://doi.org/10.1186/s12859-020-03724-x.

DPDDI: a Deep Predictor for Drug-Drug Interactions

Yue-Hua Feng[1], Shao-Wu Zhang[1*], Jian-Yu Shi[2*]
1 Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China
2 School of Life Sciences, Northwestern Polytechnical University, Xi'an, China
* Corresponding authors. Emails: zhangsw@nwpu.edu.cn; jianyushi@nwpu.edu.cn

Abstract

Background: The treatment of complex diseases taking multiple drugs becomes popular. However, drug-drug interactions (DDIs) may give rise to the risk of unanticipated adverse effects and even unknown toxicity. DDI detection in the wet lab is expensive and time-consuming. Thus, it is highly desired to develop the computational methods for predicting DDIs. Generally, most of the existing computational methods predict DDIs by extracting the chemical and biological features of drugs from diverse drug-related properties, however some drug properties are costly obtained and not available in many cases.

Results: In this work, we presented a novel method (called DPDDI) to predict DDIs by extracting the network structure features of drugs from DDI network with graph convolution network (GCN) and constructing the deep neural network (DNN) model as a predictor. GCN learns the low-dimensional feature representations of drugs for capturing the topological relationship to their neighborhood of drugs in DDI network. DNN predictor concatenates the latent feature vectors of any two drugs as the feature vector of the corresponding drug pairs to train a DNN for predicting the potential drug-drug interactions. The experiment results show that our DPDDI outperforms other four state-of-the-art methods; the GCN-derived latent features greatly outperform other features derived from chemical, biological or anatomical properties of drugs; the concatenation feature aggregation operator is better than other two feature aggregation operators (i.e., inner product and summation). The results in case studies indicates that DPDDI has the good capability for predicting the new DDIs.

Conclusion—We proposed an effective and robust method of DPDDI to predict the

potential DDIs, which just utilizes the DDI network information, working well without drug properties (i.e., drug chemical and biological properties). It can be expected that DPDDI can be helpful in other DDI-related scenarios, such as the detection of unexpected side effects, and the guidance of drug combination.

Keywords—Drug-drug interaction, DDI prediction, graph convolution network (GCN), feature extraction, deep neural network

Background

Due to synergistic effects caused by drug-drug interactions (DDIs), the combinational treatment of multiple drugs for complex diseases are popular nowadays [1]. However, unexpected DDI can also trigger side effects, adverse reactions, and even serious toxicity. They lead to patients in danger [2]. As the need of multi-drug treatments is increasing, the identification of DDIs is urgent. Nevertheless, it is expensive and time-consuming to detect DDIs among a large scale of drug pairs both in vitro and in vivo. To assist to screen DDIs, computational approaches have been developed to deduce candidate drug-drug interactions.

Existing computational methods can be roughly classified into two categories: text mining-based and machine learning-based methods. The text mining-based methods discover and collect annotated DDIs from scientific literature, electronic medical records[3, 4], insurance claim databases, and the FDA Adverse Event Reporting System[5]. They are very useful in building DDI-related databases. However, those methods cannot detect unannotated DDIs, and they cannot give an alert to potential DDIs before a combinational treatment is made [2]. In contrast, machine learning-based methods provide a promising way to identify unannotated potential drug-drug interactions for downstream experimental validations.

Usually, machine learning-based methods consist of the feature extractor and the supervised predictor. The feature extractor represents drugs in a form of feature vector according to drug properties, such as chemical structure[2, 6-14], targets[2, 8-11],

Anatomical Therapeutic Chemical classification (ATC) codes [8-10, 12], side effects [8, 9, 11, 13, 14], medication and/or clinical observations[11]. However, some drug properties can be not obtained in many cases.

The supervised predictor is usually implemented by classifier algorithms, such as KNN[12], SVM[12], logistic regression[2, 8, 10], decision tree[10], naïve Bayes[10]), and network propagation methods, such as reasoning over drug-drug network structure[6-8], label propagation[13], random walk[11, 15], probabilistic soft logic[9, 10]) or matrix factorization[14]. Usually, the predictor first trains a model with both feature vectors/similarity matrices and annotated DDI labels, then deduces potential DDIs with the well-trained model. Most methods utilize a single predictor [2, 5-8, 13-16], while some of them integrate multiple predictors to obtain the final prediction [10, 12].

In general, the predicting performance of existing approaches heavily relies on the quality of handcrafted features derived from the drug properties. However, some drug properties may not always be available. One common solution is to remove these drugs that lack a certain drug property, which results in small-scale pruned datasets and thus is not pragmatic and suitable in the real scenario[17]. In addition, some handcrafted drug features may not be precise enough to represent or characterize drug, which may fail to help build a robust and accurate model for link prediction.

As one of the most popular graph embedding methods, Graph Convolution Network (GCN) provides a promising way to predict DDIs in the condition of missing some properties of drugs. Inspired by the traditional convolutional neural networks (CNNs) operating on regular Euclidean data like images (2D grid) and text (1D sequence)[18], GCN formulates convolution on an irregular graph in non-Euclidean domains, then aggregates information about each node's neighborhood to distill the network into dense vector embedding without requiring manual feature engineering[19]. The dense vector embedding, also called low-dimensional representations, are learned to preserve the structural relationships

between nodes (e.g., drugs) of the network, and thus can be used as features in building machine learning models for various downstream tasks, such as link prediction[17]. Recently, the GCN has been applied to the field of drug development and discovery[20], such as molecular activity prediction[21], drug side effect prediction[22], drug target interactions prediction[23].

In this work, we introduced a deep predictor for predicting drug-drug interactions (called DPDDI) by using a graph convolution network (GCN) to learn a low-dimensional feature representation for each drug in the DDI networks, and adopting the deep neural network (DNN) to train models. GCN characterizes drugs in a graph embedding space for capturing the topological relationship to their neighborhood drugs. The experimental results demonstrate that our DPDDI can effectively predict DDIs, and outperforms other existing state-of-art methods.

Results

In this section, we first introduce how to set the parameters of DPDDI predictor, then compare the performance of DPDDI with other four state-of-the-art methods in 5-fold cross-validation (5CV) test. We also compare the results of three feature aggregation operators, discussing the effect of negative sample sampling rate, and the robustness of DPDDI on different scale dataset. In the end, we show the effectiveness of DPDDI through a case study.

In statistical prediction, the jackknife test and q-fold cross-validation (CV) test are often used to examine the effectiveness of a predictor[24]. Of the two test methods, the jackknife test is deemed the least arbitrary that can always yield a unique result [25]. However, for large scale database, the jackknife test needs to spend lots of time to generate the prediction results. To reduce the computational time and evaluate performance of a predictor, in this study, we adopted the 5-fold cross-validation (5CV) test as done by most investigators [26-29]. For 5CV test, the samples in the DDI dataset are randomly partitioned into 5 subsets with approximately equal size. One of the 5 subsets is singled out in turn as testing set; 90% and 10% of the other 4 subsets are used as the training samples (forming training set) and validation samples

(forming validation set), respectively. The predictor is constructed on the training set and its parameters are tuned by using the validation set. This process is repeated for 5 iterations, each time setting aside a different test subset. To avoid the bias of data randomly split, we implement 10 independent runs of 5CV, and adopt the average results to assess the performance of our DPDDI predictor.

Parameter setting

With the training DDI dataset, we performed a grid search of different parameters by seeking both the minimum value of the loss function and the best accuracy. Both the GCN-based feature extractor and the DNN-based predictor need to tune the values of the learning rate, epochs, batch size, dropout rate, as well as neuro numbers (dimensions) in hidden layers.

Specifically, with the full batch size, the GCN-based feature extractor tuned the learning rate (L-rate) from the list of {0.1, 0.01, 0.001, 0.005, 0.0001}, the Epochs from {200, 500, 800, 1000, 1200, 1400, 1600}, the Dropout from {0.01, 0.001, 0.0001}, and the hidden layer dimensions (H-dim) from {[800,512], [800,256], [800,128], [512,256], [512,128], [512,64], [256,64], [128,32]}. The DNN-based predictor tuned the learning rate (L-rate) from {0.1, 0.05, 0.01, 0.005}, the Epochs from {20, 40, 60, 80,100,140,160}, the batch size (B-size) from {10, 20, 40, 50, 60, 80}, the Dropout from {0.01, 0.001, 0.0001} and the hidden layer dimensions (H-dim) from {[128, 32], [128, 64], [64, 32], [128,64,32], [128, 32, 16], [64, 32, 16], [128, 64, 32, 16], [64, 32, 16, 4]}. The optimal values of these parameters are shown in Table 1.

Table 1. The optimal values of parameters in DPDDI

Parameters	L-rate	Epochs	Dropout	B-size	[a]I-dim	H-dim	[b]O-dim
Feature extractor	0.001	1400	0.0001	Full-batch	1562	[512,128]	128
Predictor	0.01	140	0.001	50	256	[128,64,32]	2

[a]I-dim denotes the neuro numbers in input layer; [b]O-dim denotes the neuro numbers in output layer.

Comparison with other state-of-the-art methods

To validate the effectiveness, we compared our DPDDI with other four state-of-the-art methods, including two Vilar's methods (named as Vilar 1 and Vilar 2, respectively) [6, 7], label propagation-based method (named as LP) [13] and Zhang's method (named as CE)[11] in 5-CV test. Vilar *et al* [6] integrates a Tanimoto similarity matrix of molecular structures with known DDI matrix by a linear matrix transformation to identify DDIs. Vilar *et al* [7] uses the drug interaction profile fingerprints (IPFs) to measure similarity for predicting DDIs. Label propagation method [13] applies label propagation to assign labels from known DDIs to previously unlabeled nodes by computing drug similarity-derived weights of edges on the DDI network. Zhang *et al* [11] collects a variety of drug-related data (e.g., known drug-drug interactions, drug substructures, proteins, pathways, indications, and side effects, etc.) to build many base classifiers, then develop the classifier ensemble (CE) models to predict DDIs.

To make a fair comparison, the DB2 dataset from [11] is adopted to run the prediction. In DB2 dataset, all unlabeled drug pairs are considered as the negative samples. The comparison results in 5CV test are shown in Table 2, from which we can see that DPDDI achieves the best results, and outperforms other four state-of-the-art methods across all the metrics. Specifically, DPDDI achieves the improvements of 0.2~24.9%, 6.6~64.5%, 2.2~31.5%, 2.5~50.1%, 0.6~22.1%, 8.9~50.6% against other three methods of Vilar 1, Vilar 2 and LP in terms of AUC, AUPR, Recall, Precision, Accuracy, and F1-score, respectively. Although AUC and ACC of our DPDDI are a little lower than that of Zhang's method [11], the AUPR and F_1 of DPDDI are higher than that of Zhang's method [11]. AUPR is a more significant quality measure than AUC, as it punishes much more the existence of false positive drug-drug interactions. F_1 represents the harmonic mean of precision and recall, which focus on the proportion of correctly predicted drug-drug interaction pairs. ACC focus not only on the proportion of correctly predicted drug-drug interaction pairs, but also on the proportion of correctly predicted drug-drug non-interaction pairs. For the prediction of drug-drug interaction, F_1 is more effective measure than ACC.

In addition, Zhang *et al* [11] used 9 drug-related data sources, while our DPDDI

just use the known drug-drug interaction data. If we integrate more drug-related data sources (e.g., drug substructure, drug target, drug enzyme, drug transporter, drug pathway, drug indication, drug side effect and drug off side effect used in [11]) to construct the dug-drug similarity network, using DPDDI framework to predict DDIs, DPDDI can achieve superior performance.

Table 2. Results of DPDDI and other state-of-art methods on DB2 dataset in 5CV test

Method	AUC	AUPR	Recall	Precision	ACC	F_1
Vilar 1[6]	0.707	0.262	0.495	0.253	0.719	0.334
Vilar 2[7]	0.826	0.533	0.569	0.515	0.862	0.540
LP[13]	0.851	0.799	0.685	0.729	0.809	0.706
*CE[11]	**0.957**	0.807	0.785	0.670	0.955	0.723
DPDDI	0.956	**0.907**	**0.810**	**0.754**	**0.940**	**0.840**

*The results are taken from Table 5 in Ref. [11].

Comparison of different feature aggregate operators

After obtaining the latent feature vectors of single drugs in the embedding space by GCN, we adopt three feature operators (i.e., inner product, summation and concatenation) to aggregate the feature vectors of two drugs into one feature vector for representing the drug-drug pairs. Then these aggregation feature vectors are fed into the DNN model to evaluate their effects to our DPDDI on DB1 dataset in 5CV test. As shown in Table 3, the concatenate operator achieves the best results than other two operators. Thus, we select the concatenate operator to aggregate the feature vectors of drugs in our DPDDI.

Table 3. Results of three feature aggregation operators in DPDDI on DB1 dataset in 5CV test

Operators	AUC	AUPR	Recall	Precision	ACC	F_1
Inner product	0.938	0.810	0.709	0.761	0.927	0.734
Summation	0.970	0.898	0.774	0.854	0.949	0.812
Concatenation	**0.983**	**0.925**	**0.844**	**0.836**	**0.955**	**0.840**

Comparison of drug network structure features, chemical features and biological features

In order to evaluate the effectiveness of the network structure (NS) features-based on GCN, we also used the drug chemical and biological features derived from three heterogeneous sources, such as chemical structure (CS), drug-binding proteins (DBP), and Anatomical Therapeutic Chemical Classification labels (ATC). Chemical structures of the drugs are characterized by 881-dimensional PubChem fingerprints. The DBP features of drugs are represented by 1121-dimensional binary vectors, of which each bit indicates the binding occurrence of a specific DBP across the drugs. The 118-dimensional ATC features of drugs are converted from the 7-bit ATC code by a one-hot code. These network structure features, chemical structure features, DBP features and ATC features of drugs are respectively concatenated to feed the DNN models for predicting DDIs. The results of these features with DNN on DB1 dataset in 5CV test are shown Table 4, from which we can see that the network structure features-based on GCN can generate the best performance than chemical structure, DBP and ATC features. These results indicate that the network structure features-based on GCN contains more DDI discriminant information.

Table 4. Results of the network structure features-based on GCN and other chemical and biological features with DNN on DB1 in 5CV test.

Feature	AUC	AUPR	Recall	Precision	ACC	F_1
CS	0.904	0.635	0.668	0.554	0.876	0.605
DBP	0.874	0.616	0.602	0.584	0.882	0.593
ATC	0.901	0.656	0.659	0.576	0.882	0.615
NS	**0.983**	**0.925**	**0.844**	**0.836**	**0.955**	**0.840**

Influence of dataset scale size

In order to verify the robustness of our DPDDI, we use different size datasets of DB1, DB2 and DB3 to assess the performance of DPDDI in 5CV test. DB1 dataset contains 1,562 drugs and 180,576 annotated drug-drug interactions. DB2 contains 548

drugs and 48,584 annotated drug-drug interactions. DB3 dataset contains 1,934 drugs and 230,887 annotated drug-drug interactions. As shown in Table 5, we can see that although the dataset size has some effect to the performance of DPDDI (that is, large scale size dataset can produce higher performance metrics), our DPDDI can also obtain satisfactory prediction results on the small-scale size dataset. These results show that our DPDDI have some robustness on different scale size dataset for predicting DDI.

Table 5. Results of DPDDI on different size datasets in 5CV test

Dataset	Type	Sparsity	AUC	AUPR	Recall	Precision	ACC	F_1
DB2	Small	32.4%	0.956	0.907	0.810	0.754	0.940	0.840
DB1	Medium	14.8%	0.983	0.925	0.844	0.836	0.955	0.840
DB3	Large	12.4%	0.981	0.932	0.835	0.876	0.960	0.855

By sampling different ratio samples from unlabeled drug pairs to generate the negative sample sets which are combined with the known DDI pairs (i.e., positive sample set) to form the DDI training, validation and testing datasets, we also investigate the effect of negative sample number to DPDDI. From DB1 dataset, we sample different ratio unlabeled drug pairs and combine them with the known DDI pairs to construct the datasets of DB1:1, DB1:3 and DB1:6. In DB1:1, DB1:3 and DB1:6 datasets, the ratio of positive samples (i.e., known DDI pairs) and negative samples (i.e., unlabeled drug pairs) are 1:1, 1:3 and 1:6, respectively. Figure 1 shows the results of DPDDI on DB1:1, DB1:3 and DB1:6 datasets in 5CV test. From Figure 1, we can see that DPDDI achieves the highest values in terms of AUC, AUPR, Precision, Recall, Accuracy and F_1 on DB1:1 dataset, indicating that the imbalance between positive and negative samples has great influence on the performance of DPDDI.

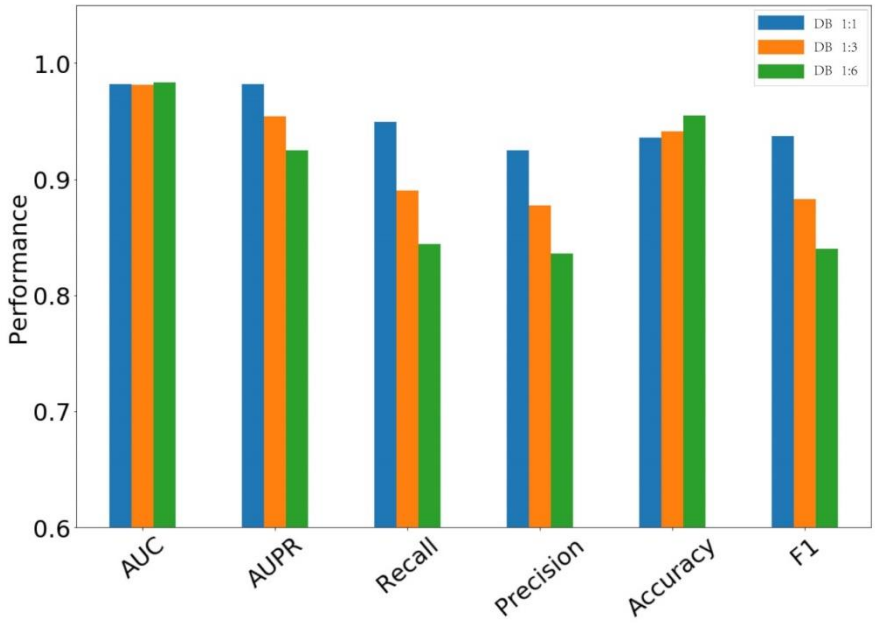

Fig. 1. Results of DPDDI on DB1:1, DB1:3 and DB1:6 datasets in 5CV test.

Case studies

In this section, we investigate the DPDDI performance of predicting the unobserved DDIs. By training DPDDI with DDI network from DB1 dataset, we will infer the possible interactions among drugs, which are not connected to each other in the DDI network. DB1 contains 180,576 annotated drug-drug interaction pairs among 1,562 drugs, and 1,038,565 unlabeled drug pairs which may contain unobserved DDIs. Higher scores of unobserved drug pairs indicate that there are higher probabilities to interact between these drugs. Table 6 shows the top 20 predicted drug-drug interactions of our DPDDI method, which are not available in DB1 dataset. By searching for the evidence of these newly predicted DDIs on DrugBank current version and Drug Interaction Checker website (Drugs.com), we find that a significant fraction of newly predicted DDIs (13 out of 20) is confirmed. For instance, the description of the interaction between drug “Doxycycline” and drug “Bleomycin” is “Doxycycline may decrease the excretion rate of Bleomycin which could result in a higher serum level”. The case studies demonstrate that our DPDDI method can effectively to detect the potential drug-drug interactions. Maybe other 7 newly predicted DDIs out of 20 are confirmed by later experiments.

In addition, among the top 20 predicted DDIs of DPDDI, we find that the drug of “doxycycline” interacts with other 8 drugs, and 5 out of 8 DDI pairs have been

confirmed by current experimental evidences. These results indicate that “doxycycline” drug may have better activity and is easy to interact with other drugs for implementing the drug efficacy.

Table 6. Top 20 predicted DDIs by DPDDI

Number	Drug 1	Drug 2	Validation source	Description
1	Doxycycline	Bleomycin	DrugBank	Doxycycline may decrease the excretion rate of Bleomycin which could result in a higher serum level.
2	Doxycycline	Rifapentine	N/A	N/A
3	Doxycycline	Fusidic acid	N/A	N/A
4	Pramipexole	Paroxetine	DrugBank	Paroxetine may increase the sedative activities of Pramipexole.
5	Luliconazole	Doxycycline	N/A	N/A
6	Netupitant	Doxycycline	DrugBank	The metabolism of Netupitant can be decreased when combined with Doxycycline.
7	Tenoxicam	Minocycline	N/A	N/A
8	Etoperidone	Tenoxicam	DrugBank	Tenoxicam may decrease the excretion rate of Etoperidone which could result in a higher serum level.
9	Pramipexole	Minocycline	N/A	N/A
10	Ropinirole	Pramipexole	DrugBank	Ropinirole may increase the sedative activities of Pramipexole.
11	Minocycline	Ropinirole	DrugBank	Minocycline may increase the central nervous system depressant (CNS depressant) activities of Ropinirole.
12	Bleomycin	Doxycycline	DrugBank	Doxycycline may decrease the excretion rate of Bleomycin which could result in a higher serum level.
13	Pramipexole	Metyrosine	drugs.com	Using metyroSINE together with pramipexole may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.
14	Osimertinib	Doxycycline	DrugBank	The metabolism of Osimertinib can be decreased when combined with Doxycycline.
15	Dronabinol	Pramipexole	DrugBank	Dronabinol may increase the sedative activities of Pramipexole.
16	Rufinamide	Tenoxicam	N/A	N/A
17	Phenobarbital	Pramipexole	drugs.com	Using PHENobarbital together with pramipexole may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.
18	Bleomycin	Mitotane	N/A	N/A
19	Fosaprepitant	Doxycycline	DrugBank	The metabolism of Fosaprepitant can be decreased when combined with Doxycycline.
20	Duloxetine	Rufinamide	drugs.com	Using DULoxetine together with rufinamide may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.

Discussions

One the key factor in DDI prediction is the feature types and feature extraction. We compared the GCN-derived DDI network structure feature with the other three chemical structure and biological features. The results in Table 4 show the great superiority of our GCN-derived DDI network structure feature across all the performance metrics. Especially, our GCN-derived DDI network structure feature achieves >20% improvement in terms of AUPR, Recall, Precision, and F_1-score. These results demonstrate that DDI network structure features-based on GCN contains more DDI discriminant information, and GCN can effectively learn a low-dimensional feature representation for each drug in the DDI network. That is, the low-dimensional representations preserve much more structural information of DDI network.

For DDI prediction of DPDDI, how to aggregate the feature vectors of two drugs into one vector for presenting one drug pair is another factor. We adopt three feature operators of inner product, summation and concatenation to aggregate the feature vectors of two drugs. The results in Table3 show that the concatenate operator achieves the best performance whereas the inner product operator gets the worst performance. Therefore, we adopt the concatenation operator as the feature aggregation in our DPDDI model.

In addition, we paid attention to how to apply negative samples in the training phase. Although many former works in other similar areas [22, 30, 31] adopted the same number of negative samples as that of positive samples to avoid the computational challenge caused by the sample imbalance. The results in Fig. 1 show that the balance sample scheme achieves the best performance in terms of AUPR, Recall, Precision and F_1 score, and the unbalance sample scheme (positive and negative sample ratio 1:6) achieves the worst performance on these terms. These results indicate that the imbalance between positive and negative samples has great influence on DPDDI. For fairly comparing with other state-of-the-art methods in which the known drug-drug interaction pairs (i.e., positive samples) and all

unlabeled drug-drug pairs（i.e., negative samples）are used to train the prediction model. Considering the factor that bigger unbalance between positive and negative samples will result in the bigger errors, we also introduce a weight W_{pos} in Eq.(2) to adjust the unbalance of negative and positive samples.

The comparison experiments (Table 1 and Table 4) with other four state-of-the-art methods and three DDI scale datasets demonstrate that DPDDI achieves superior performance, having better robustness. Investigation on the top predicted DDIs shows the competence of DPDDI for predicting the new DDIs. The superior performance of DPDDI can be attributed to the following aspects: (1) Designing a GCN model to learn the low-dimensional feature representations of drugs and capturing the structure information of DDI network. (2) Constructing a DNN model as the predictor to distinguish whether there is interaction between two drugs. DNN model can learn the non-linear relationship of drug pairs by mapping the drug pairs from a high-dimension space into a lower dimension space. Learning the low-dimensional embedding features of drugs through GCN is the main factor in our DPDDI.

Although DPDDI can predict the potential DDIs, it can only predict whether there is interaction between two drugs in DDI network. If the DDI network does not contain the drugs, DPDDI does not work for these drugs. In this condition, we can construct the drug-drug similarity network by introducing the drug chemical or biological properties, then implement our DPDDI framework to predict the novel DDIs.

Conclusions

Aiming at the preliminary screening of DDIs, this work presents a novel predicting method (namely DPDDI) by organizing a large set of DDIs as a network. DPDDI consists of a feature extractor based on graph convolution network (GCN) as well as a predictor based on deep neural network (DNN). The former characterizes drugs in a graph embedding space, where each drug is represented as a low-dimensional latent

feature vector for capturing the topological relationship to its neighborhood drugs by GCN. The latter concatenates latent feature vectors of any two drugs as the feature vector of the corresponding drug pairs to train a DNN for predicting potential interactions. Designated experiments for DPDDI bring several observations: (1) the concatenation feature aggregation operator is better than other two feature aggregation operators of the inner product and the summation; (2) the GCN-derived latent features greatly outperform other features derived from chemical, biological or anatomical properties of drugs; (3) DPDDI is robust to different scale datasets with regard to drug number, DDI number, and network sparsity; (4) the superiority of DPDDI is significantly superior to four state-of-the-art methods; (5) the finding of 13 verified DDIs out of top 20 unobserved candidates in case studies reveals the good capability of DPDDI for predicting the new DDIs. To summarize, the proposed DPDDI is an effective approach for predicting DDIs. It can be expected that DPDDI can be helpful in other DDI-related scenarios, such as the detection of unexpected side effects, and the guidance of drug combination.

Methods

Datasets

We extracted the approved small molecular drugs and their interaction relationships from DrugBank 4.0 [32] to build the DB1 dataset which contains 1,562 drugs and 180,576 annotated drug-drug interactions. In order to compare with other state-of-the-art methods, a smaller dataset (named as DB2) built by Zhang et al. [11] was adopted to evaluate the performance of our DPDDI. DB2 contains 548 drugs and 48,584 annotated drug-drug interactions. Moreover, we also collected a new and larger dataset from DrugBank 5.0 [33] to build the DB3 dataset for assess the robustness of our DPDDI, including 1,934 drugs and 230,887 annotated drug-drug interactions. In DB1, DB2 and DB3, the known drug-drug interaction pairs are used as the positive samples to build the positive set, and other unlabeled drug-drug pairs are considered as the negative samples in which we utilize the sampling strategy to build the negative set. From the point of view of the network, these three datasets can be

treated as DDI networks. The network characteristics are summarized in Table 7.

Table 7. Characteristics of DDI networks from DB1, DB2 and DB3.

Dataset	#Drug	#Interaction	#No-link	Sparsity	Max degree	Min degree
DB1	1562	180,576	1,038,565	14.8%	903	1
DB2	548	48,584	101,294	32.4%	512	1
DB3	1934	230,887	1,637,357	12.4%	1049	1

# denotes the number of drugs, link drug-drug pairs and no-link drug-drug pairs.

In order to compare our network-based features with other drug features derived from diverse drug properties, we also downloaded the drug chemical structures, Anatomical Therapeutic Chemical classification (ATC) codes and drug-binding proteins (DBPs) from DrugBank.

The chemical structure-based feature represents each drug by an 881-dimensional binary vector. Each bit in the vector indicates the occurrence or non-occurrence of a specific substructure according to Pubchem fingerprints. ATC codes are released by the World Health Organization [34], and they categorize drug substances at different levels according to organs they affect, application area, therapeutic properties, chemical and pharmacological properties. It is generally accepted that compounds with similar physicochemical properties exhibit similar biological activity. As 138 of 1,562 drugs in DB1 have no ATC code, we adopted their predicted codes by SPACE [35], which deduce ATC codes from chemical structures. To feed the 7-bit ATC code into DNN, we convert them into a one-hot code with 118 bits.

We also used drug-binding protein (DBP) data collected by [16], including 899 drug targets and 222 non-target proteins. Similarly, each drug is represented as a binary DBP-based feature vector, of which each bit indicates whether the drug binds to a specific protein.

Problem formulation

Our task is to deduce DDI candidates among those unannotated drug-drug pairs base on annotated DDIs in the form of a network. Technically, let $G(D, E)$ be a DDI network, where $D = \{d_1, d_2, \ldots, d_m\}$ is the set of m approved drugs and E denotes the interactions between them. This network can be usually represented by an $m \times m$ symmetric binary adjacent matrix $A_{m \times m} = \{a_{ij}\}$, where $a_{ij} = 1$ indicates an annotated interaction between drug d_i and drug d_j, and otherwise no annotated interaction between drug d_i and drug d_j.

DDI prediction can be solved by a three-step approach. First, function $f_1(A)$ to obtain the latent feature vector Z_i of each drug in A, where $Z_i \in R^{1 \times k} (k \ll m)$. Next, the latent vectors (Z_i and Z_j) of two drugs are aggregated to the feature vector of a drug pair. Last, function $f_2(Z_i, Z_j)$ ($Z_i, Z_j \in Z$) reconstructs the network $\hat{A}$. Function f_1 is called the feature extractor, while function f_2 is named as the predictor in our model.

In this work, by implementing the solution based on deep learning, we provide a Deep Predictor for Drug-Drug Interactions (DPDDI). DPDDI mainly consists of the following three phases: 1) Extract the low-dimensional embedding latent features of drugs from DDI network by building a GCN model; 2) Aggregate the latent feature vectors (i.e., Z_i and Z_j) of drugs d_i and d_j to represent the drug pairs; 3) Feed the fused feature vectors into a DNN to predict DDIs. The overall framework of DPDDI is illustrated in Fig.2.

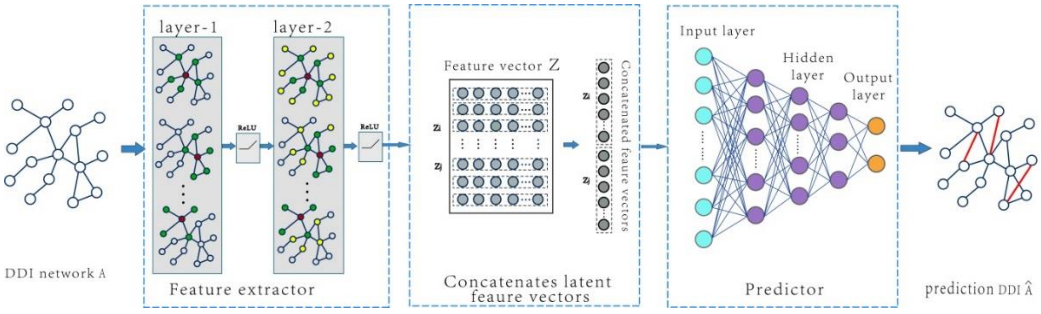

Fig. 2. The overall framework of DPDDI.

The loss of DPDDI contains two parts as follows:

$$Loss = L_f(p, q) + L_p(p, q) \quad (1)$$

where L_f is the loss of its feature extractor and L_p is the loss of its predictor. The first part adopts a binary weighted-cross-entropy as follows:

$$L_f(p,q) = -\sum_{i,j} p(a_{ij}) \log\left(q(a_{ij})\right) * W_{pos} + (1 - p(a_{ij}))(1 - \log\left(q(a_{ij})\right) \quad (2)$$

where $p(a_{ij})$ is the true label of the training interaction a_{ij}, $q(a_{ij}) = \sigma(z_i \cdot z_j^T)$ is the predicting probability computed by the inner product of latent vectors of two nodes generated by the GCN, and W_{pos} is the weight equal to the number of negative samples over the number of positive samples. The second part is defined by a binary cross-entropy as follows:

$$\mathrm{L}_p(p,q) = -\sum_{i,j} p(a_{ij}) \log\left(s(a_{ij})\right) \quad (3)$$

where $s(a_{ij})$ is the predicting probability generated by the DNN.

Feature extractor

We employ a two-layer auto-encoder of graph convolutional network(GCN)[36, 37] to obtain embedding representations of drug nodes. Each drug is represented as a latent feature vector, which contains the high-dimensional information about its neighborhood in the DDI network without manual feature engineering. Such a node embedding provides a promising way to represent the relationship between nodes in a complex network.

Technically, the GCN takes the adjacent matrix A as the input and outputs embedding vectors $\{Z_i \in R^{1\times H_p}, i = 1,2,\dots,m\}$ for every drug in the DDI network, where H_p is the dimension of the last hidden layer. Like [38] recommendation, our GCN adopts two layers as well. Suppose that $H^{(0)}$ is the feature matrix, of which each row denotes the input feature vector of each node in the network. In the case of no input feature, $H^{(0)}$ is just an identity matrix. Then, the output $H^{(1)}$ of the first hidden layer is defined as:

$$H^{(1)} = f(H^{(0)}, A) = \mathrm{ReLU}(\hat{A}H^{(0)}W^{(0)}) \quad (4)$$

where $\widehat{\mathrm{A}} = \widetilde{\mathrm{D}}^{-\frac{1}{2}}\widetilde{\mathrm{A}}\widetilde{\mathrm{D}}^{-\frac{1}{2}}$ is the symmetrically normalized adjacent matrix, $\widetilde{\mathrm{D}}_{ii} =$

$\sum_j \tilde{A}_{ij}$ and $\tilde{A} = A + I_N$, $W^{(0)} \in R^{m \times H_1}$ is the weight matrix to be learned, and ReLU is the activation function. Similarly, the output $H^{(2)}$ of the second hidden layer is recursively defined as:

$$H^{(2)} = f\left(H^{(1)}, A\right) = \mathrm{ReLU}(\hat{A}H^{(1)}W^{(1)}), \tag{5}$$

where $W^{(1)} \in R^{H_1 \times H_2}$. Because our GCN contains only two layers, $H^{(2)}$ is just the final embedding matrix $Z \in R^{m \times H_2}$.

Feature aggregation for drug pairs

So far, latent feature vector of single drug in the embedding space is obtained. The next task is to obtain feature vectors of drug pairs. Given two drugs d_i and d_j, and their latent vectors $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$ obtained by GCN, three feature operators of inner product, summation and concatenation are considered to aggregate the latent feature vectors of two drugs into a feature vector to represent the drug-drug pair. That is, we separately adopt the inner product $\boldsymbol{F}(d_i, d_j) = \boldsymbol{Z}_i \boldsymbol{Z}_j^T$, summation $\boldsymbol{F}(d_i, d_j) = \boldsymbol{Z}_i + \boldsymbol{Z}_j$ and concatenation $\boldsymbol{F}(d_i, d_j) = [\boldsymbol{Z}_i, \boldsymbol{Z}_j]$ of two drug latent vectors Z_i and Z_j to represent the drug pair (d_i, d_j).

Predictor

Given the feature vectors of drug-drug pairs, we construct a deep neural network (DNN) as the predictor in DPDDI, because of DNN proven excellent performance in classification. The predictor transforms DDI prediction as a binary classification, which is implemented by a five-layer DNN. The numbers of neurons in the layers in the DNN are 256, 128, 64, 32 and 2 respectively. ReLU is adopted as the activation function in the first four layers while SoftMax is the activation function in the last layer, which outputs how likely drug pairs are potential DDIs.

There are two steps to train our DPDDI. The first step is training a GCN for obtaining the low-dimensional embedding latent features of drugs. The parameters (i.e., learning rate, epochs, dropout, input-dim, hidden-dim, and output-dim) in GCN architecture are trained using the DDI network data. The second step is learning the

parameters (i.e., learning rate, dropout, epochs, batch-size, input-dim, hidden-dim, and output-dim) of the DNN for final DDI prediction and processing a fine-tuning for renew all parameters in the whole DPDDI framework. To explain our DPDDI method in detail, the pseudo-code is shown in Table 8.

Table 8. The pseudo-code of DPDDI

Input: DDI network A

The parameters: learning rate, epochs, dropout, batch-size, input-dim, hidden-dim, output-dim (both in Feature extractor and Predictor)

Output: DDI network $\hat{A}$ reconstructed by DPDDI

1: Initialize parameter sets $W^{(0)}$ and $W^{(1)}$ in Feature extractor.
2: Learn drug representations Z.
3: **for** epoch **in** epochs (Feature extractor in Table 1.):
4: Compute the loss function based on Eq. 2.
5: Calculate gradient and adopt Adam optimizer to update $W^{(0)}$ and $W^{(1)}$.
6: **end for**
7: Obtain the representations Z of drugs according to Eq. 4 and Eq. 5.
8: **for** each drug pair, do
9: Feature aggregation by concatenating operation.
10: **end for**
11: Initialize parameter sets in Predictor based on DNN.
12: Feed representation vector of each drug pair into Predictor.
13: **for** epoch **in** epochs (Predictor in Table 1.):
14: Compute the loss function based on Eq. 3.
15: Calculate gradient and adopt Adam optimizer to update parameter sets.
16: **end for**
17: Obtain the DDI network $\hat{A}$.

Evaluation metrics

The following metrics of accuracy (ACC), Recall, Precision and F_1-score are used to measure the performance of DPDDI.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

where TP and TN are the number of correctly predicted DDI pairs and unlabeled drug-drug pairs, respectively; FP and FN are the number of incorrectly predicted DDI pairs and unlabeled drug-drug pairs, respectively.

We also used the metrics of AUC and AUPR to measure the performance of our DPDDI. AUC is the area under the receiver operating characteristic (ROC) curve which plots the true-positive rate (i.e., $TP/(TP+FN)$) versus the false-positive rate (i.e., $FP/(FP+TN)$) at different cutoffs. AUPR is the area under the precision–recall curve which plots the ratio of true positives among all positive predictions for each given recall rate.

Abbreviations

DDIs: drug-drug interactions

GCN: graph convolution network

DNN: deep neural network

ATC: Anatomical Therapeutic Chemical classification

DBP: drug-binding protein

AUC: Area Under the receiver operating characteristic Curve

AUPR: Area Under the Precision-Recall curve

ACC: accuracy

References

1. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MCJNB: **Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions**. *Nat Biotechnol* 2017, **35**:463-474.
2. Takeda T, Hao M, Cheng T, Bryant SH, Wang YJJoC: **Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge**. *J Cheminform* 2017, **9**(16).
3. Pathak J, Kiefer RC, Chute CG: **Using linked data for mining drug-drug interactions in electronic health records**. *Stud Health Technol Inform* 2013, **192**:682-686.
4. Duke JD, Xu H, Zhiping W, Abhinita S, D. KS, Xiaochun L, D. HS, Yan J, Thomas CJ, Biology OMJJPC: **Literature Based Drug Interaction Prediction with Clinical Assessment Using Electronic Medical Records: Novel Myopathy Associated Drug Interactions**. *PloS cOMPUT biiol* 2012, **8**(8):e1002614.
5. Vilar S, Friedman C, Hripcsak GJBiB: **Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media**. *Brief Bioinform* 2018, **19**(5):863-877.

6. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman CJJotAMIAJ: **Drug-drug interaction through molecular structure similarity analysis**. *Journal of the American Meidical Informatics Association* 2012, **19**(6):1066-1074.
7. Vilar S, Uriarte E, Santana L, Tatonetti NP, Friedman CJPo: **Detection of drug-drug interactions by modeling interaction profile fingerprints**. *PLOS ONE* 2013, **8**(3):e58321.
8. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan RJMSB: **INDI: a computational framework for inferring drug interactions and their associated recommendations**. *Mol Syst Biol* 2012, **8**(1).
9. Sridhar D, Fakhraei S, Getoor LJB: **A Probabilistic Approach for Collective Similarity-based Drug-Drug Interaction Prediction**. *Bioinformatics* 2016, **32**(20):3175-3182.
10. Cheng F, Zhao Z: **Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties**. *Journal of the American Meidical Informatics Association* 2014, **21**(e2):e278-e286.
11. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li XJ: **Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data**. *BMC bioinformatics* 2017, **18**(1):18.
12. Andrej K, Polonca F, Brane Le: **Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning**. *Plos One* 2018, **13**(5):e0196865-.
13. Zhang P, Wang F, Hu J, Sorrentino R: **Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects**. *Scientific Reports* 2015, **5**(1):12339.
14. Yu H, Mao K, Shi J, Huang H, Chen Z, Dong K: **Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization**. *BMC systerms biology* 2018, **12**(1):14.
15. Park K, Kim D, Ha S, Lee: **Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks**. *PLOS ONE* 2015, **10**(10):e0140816.
16. Shi J-Y, Mao K-T, Yu H, Yiu S: **Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization**. *Journal of cheminformatics* 2019, **11**(1).
17. Yue X, Wang Z, Huang J, Parthasarathy S, Moosavinasab S, Huang Y, Lin SM, Zhang W, Zhang P, Sun H: **Graph embedding on biomedical networks: methods, applications and evaluations**. *Bioinformatics* 2020, **36**(4):1241-1251.
18. Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M: **Graph Neural Networks: A Review of Methods and Applications**. In: *arXiv e-prints arXiv:181208434.* 2018.
19. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS: **A Comprehensive Survey on Graph Neural Networks**. In: *In: arXiv e-prints arXiv:190100596.* 2020/03/29 edn; 2020.
20. Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F: **Graph convolutional networks for computational drug development and discovery**. *Brief Bioinform* 2020, **21**(3):919-935.
21. Pham T, Tran T, Venkatesh S: **Graph Memory Networks for Molecular Activity Prediction**. In: *arXiv e-prints arXiv:180102622.* 2018.
22. Zitnik M, Agrawal M, Leskovec: **Modeling polypharmacy side effects with graph convolutional networks**. *Bioinformatics* 2018, **34**(13):i457-i466.
23. Gao KY, Fokoue A, Luo H, Iyengar A, Zhang P: **Interpretable Drug Target Prediction Using Deep Neural Representation**. In: *Twenty-Seventh International Joint Conference on Artificial*

Intelligence IJCAI-18: 7 2018. 3371--3377.

24. Chou KC, Zhang CT: **Prediction of Protein Structural Classes**. 1995, **30**(4):275-349.
25. Chou KC: **Some remarks on protein attribute prediction and pseudo amino acid composition**. *J Theor Biol* 2011, **273**(1):236-247.
26. Fan X-N, Zhang S-W: **LPI-BLS: Predicting lncRNA-protein interactions with a broad learning system-based stacked ensemble classifier**. *Neurocomputing* 2019, **370**:88-93.
27. Yan XY, Zhang SW: **Identifying Drug-Target Interactions with Decision Templates**. *Curr Protein Pept Sci*. 2018, **19**(5): 498-506.
28. Zhang Y, Qiu Y, Cui Y, Liu S, Zhang W: **Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning**. *Methods* 2020.
29. Zheng Y, Peng H, Zhang X, Zhao Z: **DDI-PULearn: a positive-unlabeled learning method for large-scale prediction of drug-drug interactions**. Bioinformatics .2019, **20**(Suppl 19):661.
30. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J: **Distributed representations of words and phrases and their compositionality**. In: *Advances in neural information processing systems: 2013*. 3111-3119.
31. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G: **Complex Embeddings for Simple Link Prediction**. In: *arXiv e-prints arXiv:160606357.* 2017.
32. Vivian L, Craig K, Yannick D, Tim J, Guo AC, Liu Y, Adam M, David A, Michael W, Vanessa NJNAR: **DrugBank 4.0: shedding new light on drug metabolism**. *Nucleic acids research* 2013, **42**(D1):D1091-D1097.
33. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z *et al*: **DrugBank 5.0: a major update to the DrugBank database for 2018**. *Nucleic Acids Research* 2017, **46**(D1):D1074-D1082.
34. Skrbo A, Begović B, Skrbo S: **Classification of drugs using the ATC system (Anatomic, Therapeutic, Chemical Classification) and the latest changes**. *Medicinski arhiv* 2004, **58**(1 Suppl 2):138-141.
35. Liu Z, Guo F, Gu J, Wang Y, Li Y, Wang D, Lu L, Li D, He F: **Similarity-based prediction for Anatomical Therapeutic Chemical classification of drugs by integrating multiple data sources**. *Bioinformatics* 2015, **31**(11):1788-1795.
36. Kipf TN, Welling M: **Variational Graph Auto-Encoders**. In: *arXiv e-prints arXiv:161107308.* 2016.
37. Kipf TN, Welling MJapa: **Semi-supervised classification with graph convolutional networks**. 2016.
38. Defferrard M, Bresson X, Vandergheynst P: **Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering**. *Advances in neural information processing systems* 2016, **29**:3844-3852.

Acknowledgments

We acknowledge anonymous reviewers for the valuable comments on the original manuscript.

Funding

This work has been supported by the National Natural Science Foundation of China (No. 61873202, PI：SWZ and No. 61872297, PI: JYS) and Shaanxi Provincial Key R&D Program, China (NO. 2020KW-063, PI: JYS).

Availability of data and materials

All the code and data are openly available at the website of https://github.com/NWPU-903PR/DPDDI.

Author information

Affiliations

Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China
Yue-Hua Feng, Shao-Wu Zhang
School of Life Sciences, Northwestern Polytechnical University, Xi'an, China
Jian-Yu Shi

Authors' contributions

YHF collected the dataset, performed the experiments and drafted the manuscript. JYS analyzed the result. Both JYS and SWZ modified manuscript and they are the corresponding authors. All authors read and approved the final manuscript.

Correspondence

Shao-Wu Zhang, zhangsw@nwpu.edu.cn ; Jian-Yu Shi, jianyushi@nwpu.edu.cn;

Ethics declarations

Ethics approval and consent to participate

No ethics approval was required for the study.

Consent for publication

Not applicable.

Competing interests

None of the authors has any competing interests.

Figure Legends

Fig. 1. Results of DPDDI on DB1:1, DB1:3 and DB1:6 datasets in 5CV test.
According to the different ratio of positive and negative samples (including DB1:1, DB1:3 and DB1:6 on DB1), different performance are obtained.

Fig. 2. The overall framework of DPDDI.

The main steps are as follows. First, the feature extractor of DPDDI constructs a two-layer graph convolutional network (GCN) to obtain drug latent features, which capture the complex relations between the drug nodes in the DDI network. Then, each pair of drugs is represented as a feature vector by concatenating the corresponding latent features of the drugs. Last, the feature vectors of representing the drug pairs are fed into a deep neural network to train the predictor to deduce potential DDIs.

Figures

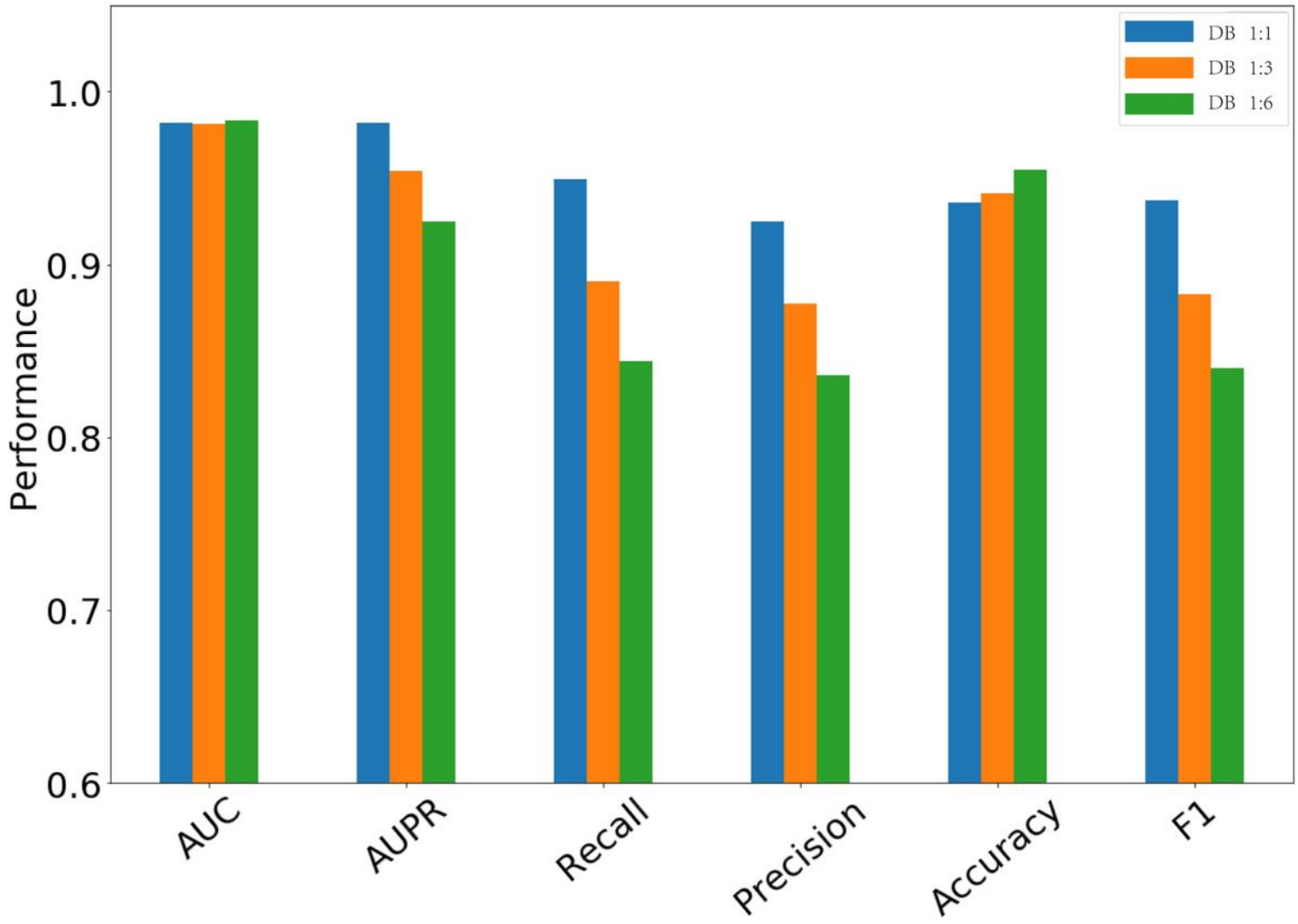

Figure 1

Results of DPDDI on DB1:1, DB1:3 and DB1:6 datasets in 5CV test. According to the different ratio of positive and negative samples (including DB1:1, DB1:3 and DB1:6 on DB1), different performance are obtained.

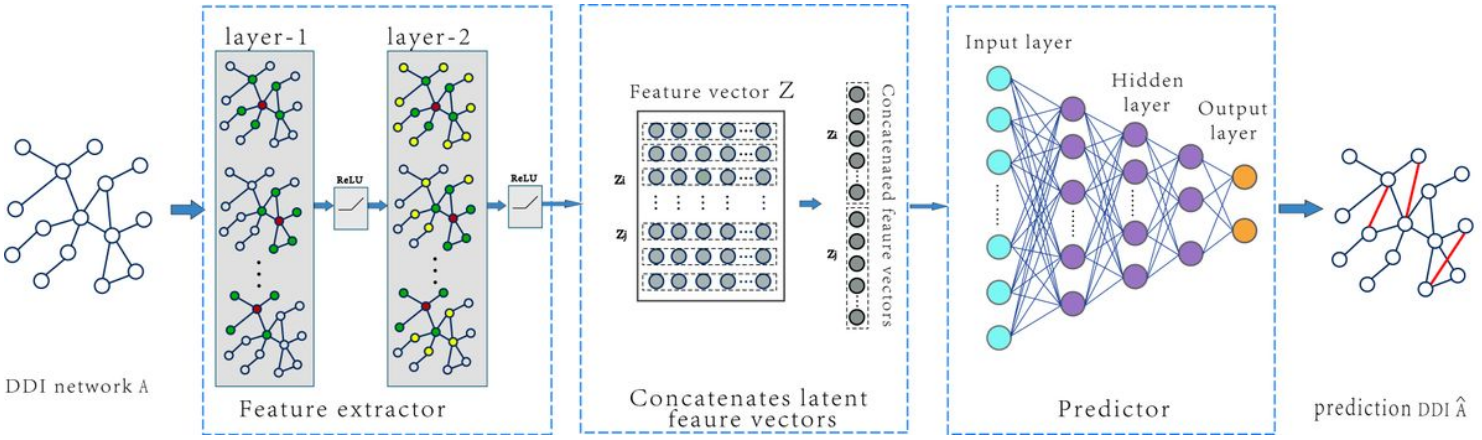

Figure 2

The overall framework of DPDDI. The main steps are as follows. First, the feature extractor of DPDDI constructs a two-layer graph convolutional network (GCN) to obtain drug latent features, which capture the complex relations between the drug nodes in the DDI network. Then, each pair of drugs is represented as a feature vector by concatenating the corresponding latent features of the drugs. Last, the feature vectors of representing the drug pairs are fed into a deep neural network to train the predictor to deduce potential DDIs.