

MTDDI: a graph convolutional network framework for predicting Multi-Type Drug-Drug Interactions

YueHua Feng (✉ fengyuehua@mail.nwpu.edu.cn)

Northwestern Polytechnical University <https://orcid.org/0000-0002-3783-1305>

Shao-Wu Zhang

Northwestern Polytechnical University <https://orcid.org/0000-0003-1305-7447>

Qing-Qing Zhang

Northwestern Polytechnical University <https://orcid.org/0000-0002-7931-1834>

Chu-Han Zhang

Northwestern Polytechnical University <https://orcid.org/0000-0002-2897-3918>

Jian-Yu Shi

Northwestern Polytechnical University <https://orcid.org/0000-0002-2303-273X>

Research article

Keywords: Drug-drug interactions (DDIs), multi-type DDIs prediction, graph convolution network (GCN), tensor factorization, deep neural network, multiple relation prediction, similarity regularization

Posted Date: April 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-397281/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

MTDDI: a graph convolutional network framework for predicting Multi-Type Drug-Drug Interactions

Yue-Hua Feng¹, Shao-Wu Zhang^{1*}, Qing-Qing Zhang¹, Chu-Han Zhang², Jian-Yu Shi^{3*}

¹ Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

² School of Software, Northwestern Polytechnical University, Xian, 710072, China

³ School of Life Sciences, Northwestern Polytechnical University, Xi'an, 710072, China

* Correspondence: zhangsw@nwpu.edu.cn ; jianyushi@nwpu.edu.cn

Abstract— Although the polypharmacy has both higher therapeutic efficacy and less drug resistance in combating complex diseases, drug-drug interactions (DDIs) may trigger unexpected pharmacological effects, such as side effects, adverse reactions, or even serious toxicity. Thus, it is crucial to identify DDIs and explore its underlying mechanism (e.g., DDI types) for polypharmacy safety. However, the detection of DDIs in assays is still time-consuming and costly, due to the need of experimental search over a large drug combinational space. Machine learning methods have been proved as a promising and efficient method for preliminary DDI screening. Most shallow learning-based predictive methods focus on whether a drug interacts with another or not. Although deep learning (DL)-based predictive methods address a more realistic screening task for identifying the DDI types, they only predict the DDI types of known DDI, ignoring the structural relationship between DDI entries, and they also cannot reveal the knowledge about the dependence between DDI types. Thus, here we proposed a novel end-to-end deep learning-based predictive method (called MTDDI) to predict DDIs as well as its types, exploring the underlying mechanism of DDIs. MTDDI designs an encoder derived from enhanced deep relational graph convolutional networks to capture the structural relationship between multi-type DDI entries, and adopts the tensor-like decoder to uniformly model both single-fold interactions and multi-fold interactions to reflect the relation between DDI types. The results show that our MTDDI is superior to other state-of-the-art deep learning-based methods. For predicting the multi-type DDIs with unknown DDIs in case of both

single-fold DDIs and multi-fold DDIs, we validated the effectiveness and the practical capability of our MTDDI. More importantly, MTDDI can reveal the dependency between DDI types. These crucial observations are beneficial to uncover the mechanism and regularity of DDIs.

Keywords—Drug-drug interactions (DDIs), multi-type DDIs prediction, graph convolution network (GCN), tensor factorization, deep neural network, multiple relation prediction, similarity regularization

1. Introduction

The polypharmacy, also termed as drug combination, is becoming a promising strategy for treating complex diseases (e.g., diabetes and cancer) in recent years[1]. When two or more drugs are taken together, they may trigger unexpected side effects, adverse reactions, and even serious toxicity [2]. The pharmacological effects triggered by multiple drugs in the treatment are named drug-drug interactions (DDIs). DDIs can be divided into two cases. One case is that a pair of drugs triggers only one pharmacological effect, another is that a pair of drugs causes two or more related pharmacological effects. We call the former as a single-fold interaction and the latter as a multiple-fold interaction. For example, the interaction between Sucralfate and Metoclopramide tells “Sucralfate may decrease the excretion rate of Metoclopramide, resulting in a higher serum level”. Apparently, the pair of these two drugs may trigger two related pharmacokinetic effects, that is, Excretion and Serum Concentration. Therefore, it is crucial to identify DDIs and unravel their underlying mechanisms for polypharmacy safety. However, it is still both time-consuming and costly to detect DDIs among a large scale of drug pairs in assays. Over the past decade, this build-up of experimentally-determined DDI entries boosts the application of computational methods to find the potential DDIs [3], especially machine learning-based methods.

Various machine learning methods have been proved as a promising method to provide a preliminary screening of DDIs for further experimental validation with the

advantages of both high efficiency and low costs. Generally, the machine learning-based methods [4-15] use the approved DDIs training the predictive models to infer the potential DDIs among massive unlabeled drug pairs by extracting the drug features from diverse drug property source, such as chemical structure [4, 6-9], targets [4-7], anatomical taxonomy [5, 8, 10] and phenotypic observation [5, 7, 9, 10], or extracting the drug similarity features [5, 6, 9, 10, 16-18], or training the deep learning models to extract better features from raw features [19-21]. However, most of these existing methods focus on whether a drug interacts with another or not.

Recent years, other deep learning(DL)-based predictive methods [11, 12, 22] have been developed to address another screening task of identifying the pharmacological effects caused by known DDIs, that is, predicting multi-type DDIs. For example, DeepDDI [11] designs a nine-layer deep neural network to predict 86 types of DDIs by using the structural information of drug pairs as inputs. Lee et al [22] predicts the pharmacological effects of DDIs by using three drug similarity profiles including the structural similarity profile, Gene Ontology term similarity profiles and target gene similarity profiles of known drug pairs to train the three-layer autoencoder and a eight-layer deep feed-forward network. DDIMDL [12] predicts DDI events by using the drug similarity features computed from chemical substructures, targets, enzymes and pathways to separately train three-layer deep neural networks (DNNs), and then averages (sums up) individual predictions of those trained DNNs as the final prediction.

Despite these efforts on identifying multi-type DDIs, there still exists the following space to improvement for DL-based methods. i) Existing DL-based methods require the known DDI as input, while the interactions of most drug pairs are unknown. Therefore, it is necessary to develop new algorithms to identify whether an unknown drug pair has one or more pharmacological effects. ii) More DDIs can form an interaction network that help to improve the predictor's performance, however existing DL-based methods treat drug pairs as independent samples, ignoring the structural relationship between DDI entries. iii) Existing DL-based methods cannot

reveal the knowledge (e.g., the excretion of a drug slows down due to its increasing serum concentration caused by a DDI) about the relation between DDI types. To address above issues, we proposed a novel predictive method (called MTDDI) to identify whether an unknown drug pair results in one or more pharmacological effects. The main contributions of our work are described as follows: i) MTDDI leverages an encoder by an enhanced relational graph convolutional network (R-GCN) to capture the structural relationship between multiple-type DDI entries. ii) MTDDI employs a tensor-like decoder to uniformly model both single-fold interactions and multi-fold interactions for identifying whether an unlabeled type-specific drug pair results in one or more pharmacological effects. iii) MTDDI adopts a set of type-specific feature importance matrices (i.e., a tensor) in decoder to reveal the dependency between DDI types by calculating their correlations.

2. Materials and Methods

2.1. Datasets

We built the multi-type DDI dataset by collecting DDI entries from DrugBank (July 16, 2020[23]) in the following steps. First, we downloaded the completed XML-formatted database (including the comprehensive profiles of 11,440 drugs), from which we selected 2,926 small-molecule drugs and their drug chemical structures and drug binding proteins. After extracting all descriptive sentences of DDIs from the XML file, we totally collected 859,662 interaction entries among 2,926 drugs. Furthermore, we obtained 274 different interaction patterns by parsing these sentences. According to pharmacological effects triggered by DDIs [24], we finally grouped these patterns into 11 types of DDIs, including Absorption, Metabolism, Serum Concentration, Excretion, Synergy Activity, Antagonism Activity, Toxicity Activity, Adverse Effect, Antagonism Effect, Synergy Effect, and PD triggered by PK [25].

2.2. Problem formulation

Suppose n drugs $\mathcal{D} = \{d_i\}$ and k interactions $\mathcal{L} = \{l_{ij}\}$ among them. The

traditional DDI prediction, multi-type DDI classification and multi-type DDI prediction are different pharmacological tasks.

- The task of traditional DDI prediction learns a function mapping $\mathcal{F}: \mathcal{D} \times \mathcal{D} \rightarrow \{0,1\}$ to deduce potential interactions between unlabeled drug pairs among \mathcal{D} (Fig.1-A).
- The task of multi-type DDI classification identifies what pharmacological effects caused by known DDIs are (Fig.1-B). It learns a function mapping $\mathcal{F}: \mathcal{L} \rightarrow \{t_i\}, i = 1, 2 \dots T$, where t_i is the pharmacological effect type of DDIs, and T is the cardinality of all pharmacological effects.
- The task of multi-type DDI prediction directly discriminates whether an unknown drug pair results in one or more pharmacological effects of interest (Fig.1-C). It learns a set of functions mapping $\mathcal{F}_i: \mathcal{D} \times \mathcal{D} \rightarrow [0,1], i = 1, 2 \dots T$.

This work focuses on the task of multi-type DDI prediction since the second task is just its degraded version. Referring to DDI-triggered pharmacological effects as interaction types, we represent a set of multi-type DDIs as a multi-relation complex network $G(\mathbf{V}, \mathbf{R})$, where vertices are drugs and edges between vertices are multi-type interactions (Fig.2-A). Let $\mathbf{V} = \{v_1, v_2, \dots, v_n\}$ be the vertex set, $\mathbf{T} = \{t_1, t_2, \dots, t_m\}$ be the interaction type set, and (v_i, t_r, v_j) be the interaction of type r caused by the pair of drug d_i and drug d_j . Furthermore, G is decomposed into m sliced sub-networks $G = \{G_1, G_2, \dots, G_m\}$ regarding interaction types (Fig.2-A). Each slice, denoted as G_r , is represented a symmetric adjacent matrix $\mathbf{A}_{n \times n}^r = \{a_{ij}^r\}, i, j = 1, 2, \dots, n$, where $a_{ij}^r = 1$ indicates an approved interaction of type r between drug d_i and drug d_j , and $a_{ij}^r = 0$ otherwise. These binary adjacent matrices naturally form a 3-order multi-relational tensor $\mathbf{R} \in \mathbb{R}^{n \times n \times m}$ (Fig.2-A). Besides, pairwise similarities of all drugs among \mathcal{D} are organized into a similarity matrix, $\mathbf{S} = \{s_{u,v}\} \in$

$[0,1], u, v = 1, 2, \dots, n.$

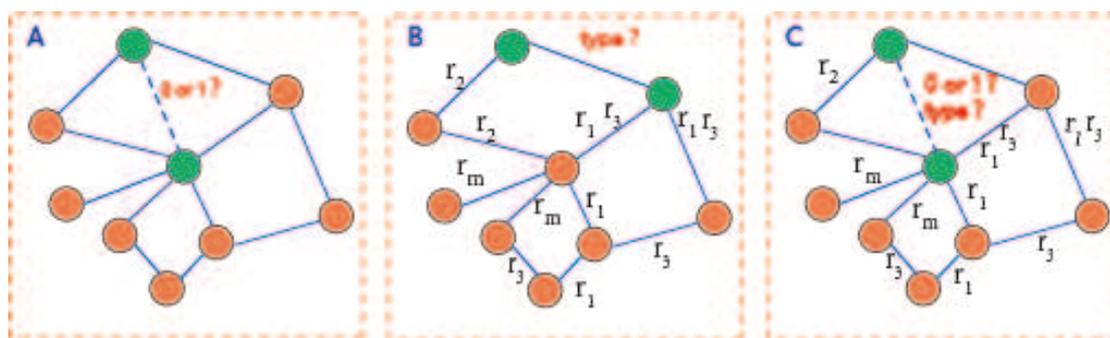


Figure 1. Three tasks in DDIs prediction: (A) Binary DDI prediction. (b) Multi-type DDI classification. (C) Multi-type DDI prediction.

2.3 Feature extraction

In addition to interaction entries, we extracted drug chemical structures, which are represented by SMILES strings, as well as drug binding proteins (DBPs), including targets, enzymes, transporters, and carriers. Drug chemical structures were encoded into feature vectors by Extended Connectivity Fingerprints (ECFPs) [26] and MACCSkeys Fingerprints [26], respectively. ECFPs represent a molecular structure through circular atom neighborhoods as 1024-dimensional binary vector, where each element denotes the presence or the absence of a specific functional substructures. In contrast, the MACCSkeys Fingerprints represent a molecular structure as 166-dimensional binary vector w.r.t. a set of pre-defined substructures. These two fingerprints are computed by the rdkit package of python, and the radius of ECFPs neighborhood is set to 4.

Moreover, we consider DBPs (targets, transporters, enzymes, and carrier proteins) as the third type of drug features, because they are crucial factors when a DDI occurs. Sequentially, drug is represented as a 3334-dimensional binary vector in which each element indicates whether the drug binds to a specific protein. Finally, by calculating Tanimoto coefficients between drug feature vectors, we obtained three drug similarities derived from ECFP_4, MACCSkeys and DBPs, respectively.

2.4 Model construction

Upon above representation of multi-type DDIs, we cast the task of multi-type DDI prediction as the multi-relational link prediction, and design an end-to-end Multiple-Type Predictor for Drug-Drug Interactions (MTDDI) to address this task. MTDDI contains an encoder \mathcal{F}_e and a decoder \mathcal{F}_d .

Derived from the multi-relation GCN (R-GCN) [27-30], we construct a multi-layer R-GCN in which encoder \mathcal{F}_e extracts a global latent feature matrix $\mathbf{Z}_{n \times k}$ ($k \ll n$) by capturing the topological feature matrices $\{\mathbf{Z}_{n \times k}^r\}$ of all drugs across $\{G_r\}$. However, the primary multi-layer GCN causes the over-smoothing issue that makes all the nodes in a network have highly similar feature values. To relax the over-smoothing issue, \mathcal{F}_e doesn't use the outputting embedding representations of its final layer, but it sums the embedding representations (named residuals) of its hidden layers together as its final embedding feature matrix \mathbf{Z} . In addition, considering a few of possible missing interactions among the network, \mathcal{F}_e utilizes a pre-defined drug similarity matrix to constrain the similar drugs more close in the embedding space.

Since the original decoder in the primary GCN [30] is just an inner production $\mathbf{Z}\mathbf{Z}^T$ between drug embedding vectors, it cannot reflect the essence of multi-type interactions. R-GCN employs RESCAL [31], which utilizes m additional type-specific feature association matrices \mathbf{M}_r to capture the essence of multi-type interactions (i.e., $\mathbf{Z}\mathbf{M}_r\mathbf{Z}^T$). Inspired by literature[27, 32], we suppose that feature importance varies across interaction types, and we also assume that interaction types are not completely independent to each other. Therefore, our decoder \mathcal{F}_d adopts a tensor factorization-like matrix operation to integrate the embedding feature matrix \mathbf{Z} , m type-specific feature importance matrices, and an average feature association matrix to reconstruct the multi-type DDIs network (i.e., $\mathbf{Z}\mathbf{D}_r\mathbf{R}\mathbf{D}_r\mathbf{Z}^T$).

Finally, our MTDDI trains \mathcal{F}_e and \mathcal{F}_d simultaneously to obtain an end-to-end model for implementing the multi-type DDI prediction.

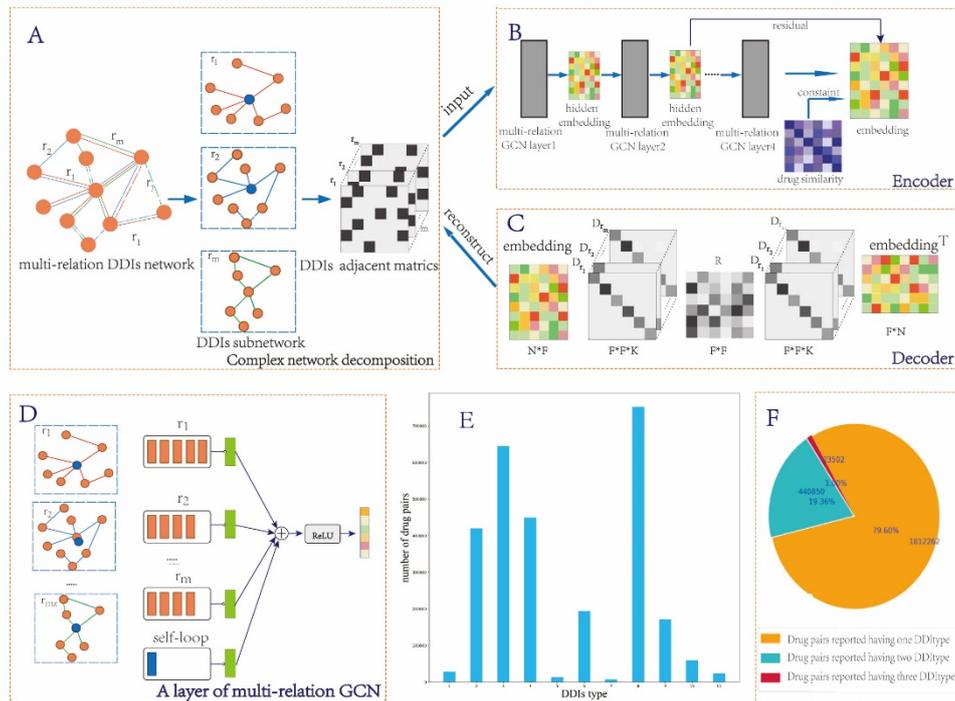


Figure 2. Overall framework of MTDDI and multi-type DDI statistics. **(A)** Decomposition of the multi-type DDIs network. The multi-type DDI network is decomposed into m sliced (i.e., type number) subnetworks, which are represented by m adjacent matrices and are taken as the input of the encoder. **(B)** The encoder of MTDDI. It constructs a p -layer multi-relation GCN (R-GCN) to encode drugs in the multi-type DDI network into embedding vectors (i.e., rows in the colorful matrix) by capturing their complex topological properties. A residual strategy (i.e., the black arrow) is added from the second hidden layer to the last hidden layer. Meanwhile, a drug similarity matrix is employed to constrain similar drugs as close as possible in the embedding space (i.e., the purple matrix). **(C)** The decoder of MTDDI. It is a tensor factorization-like matrix operation, which integrates the embedding feature matrix, type-specific feature importance matrices $\{D_r\}$, and an average feature association matrix R to reconstruct the multi-type DDIs network. **(D)** An example to illustrate a layer of R-GCN in the encoder. An interest node (i.e., blue node) aggregates both the features of its first-order neighbor nodes (i.e., orange) and its own in each of m sliced networks to update its features (i.e., green bar). Then, all the updated features are accumulated and passed through a ReLU activation function to produce its final embedding (i.e., the colorful vector). The whole multi-type DDI network are propagated by a p -layer R-GCN to capture the information of its p -th order neighbors. **(E)** Statistics on different pharmacological effects caused by DDIs. From the left to the right, the interaction types are: Absorption, Metabolism, Serum Concentration, Excretion, Synergy Activity, Antagonism Activity, Toxicity Activity, Adverse Effect, Antagonism Effect, Synergy Effect, and PD triggered by PK. Y-axis indicates their occurring numbers. **(F)** Proportional distribution of the number of

single-fold and multi-fold DDIs. 79.6% DDIs are single-fold, 19.36% are two-fold and 1.04% are three-fold.

2.4.1 Encoder in Multi-relation graph convolutional network

We employed the extension GCN (i.e., R-GCN) to extract the node embedding in the multi-type DDI network. First, the network G is decomposed into m sliced sub-networks $\{G_1, G_2, \dots, G_m\}$, in which each slice accounts for a specific interaction type (Fig.2-A). Then, both the feature vector $\mathbf{h}_i^{(0)}$ of drug d_i (or node v_i) and those of its neighbors in G_r are aggregated by a graph convolutional operation. After that, similar aggregations across all the sliced subnetworks are further summed up to generate the updated feature vector $\mathbf{h}_i^{(1)}$ of drug d_i . Such a single layer of R-GCN integrates the topological neighborhood of drug d_i across interaction types which it involves. For any layer in a multi-layer R-GCN, the general propagation rule is defined as:

$$h_i^{(k+1)} = \sigma \left(\sum_r \left(\sum_{j \in N_i^r} \frac{1}{c_{i,r}} w_r^{(k)} h_j^{(k)} + w_r^{(k)} h_i^{(k)} \right) \right), \quad (1)$$

where $c_{i,r} = |N_i^r|$ is a normalization constant, N_i^r denotes the set of d_i 's neighbors in G_r , $h_i^{(k)}$ is the input feature vector and $\mathbf{w}_r^{(k)}$ is the trainable weight matrix in the k -th layer of R-GCN, and σ is a non-linear element-wise activation function (i.e., ReLU). Last, the aggregation process is propagated through p layers of R-GCN to obtain the final embedding feature vector $\mathbf{h}_i^{(p)}$ of drug d_i .

Such a multi-layer propagation of R-GCN enables the extraction of higher-order topological features of multi-type DDI network [33]. However, it usually causes the ‘over smoothing’ issue derived from GCN [33], where the features of the neighboring drugs, even all drugs in the case of many layers, are extremely similar. As a result, a good GCN contains only a few of hidden layers (e.g., the number of layers is less than or equal to 2) [28-30]. To enhance the ability of GCN’s network representation, a residual strategy is adopted to relax ‘over smoothing’ issue for multi-layer R-GCN.

Let the final embedding features outputted by the Encoder \mathcal{F}_e be \mathbf{z}_i . For a p -layer R-GCN, we set \mathbf{z}_i as:

$$\mathbf{z}_i = \sum_{k=2}^p \mathbf{h}_i^k. \quad (2)$$

Notedly, this sum requires that the dimensions in different layers are same. Due to the first hidden layer accounts for the dimension reduction of the high-dimensional one-hot features $\mathbf{h}_i^{(0)}$, the residual strategy just starts the sum from the second hidden layer.

Moreover, it is anticipated that two interacting drugs are close in the embedding space generated by \mathcal{F}_e . Thus, possible interactions can be deduced among those close drugs according to their embedding features [30]. However, the existing interaction with missing label between two drugs possibly causes their remoteness in the network. Missing interactions between these drugs would aggravate the learning of \mathcal{F}_e . Therefore, under the consideration that similar drugs tend to interact in terms of chemical structures[2] or binding proteins[34], pre-defined drug similarities, taken as a regularization item $s_{i,j} \cdot \|\mathbf{z}_i - \mathbf{z}_j\|_2^2$, is employed to constrain similar drugs as close as possible in the embedding space. Refer to Section 2.5 Loss Function for details.

2.4.2 Decoder

Once the encoder \mathcal{F}_e generates drug embedding features $\{\mathbf{z}_i\}$, which integrate topological information across interaction types, the decoder \mathcal{F}_d sequentially employs $\{\mathbf{z}_i\}$ to reconstruct the multi-type DDI network \tilde{G} . In the case of binary DDI prediction, the inner production $\mathbf{z}_i \mathbf{z}_j^T$ indicates how likely drug d_i interacts with drug d_j . In order to reflect the difference between interaction types, R-GCN employs $\mathbf{z}_i \mathbf{M}_r \mathbf{z}_j^T$ to calculate the likelihood of being a type-specific interaction, where $\{\mathbf{M}_r\}$ are m specific-type associative matrices. Inspired by literature [35], we suppose that feature importance varies across interaction types, and we also assume

that interaction types are not completely independent to each other. Therefore, our decoder \mathcal{F}_d adopts a tensor factorization-like matrix operation $\mathbf{z}_i \mathbf{D}_r \mathbf{R} \mathbf{D}_r \mathbf{z}_j^T$ to calculate the type-specific interaction likelihood. Thus, how likely the pair of drug d_i and drug d_j triggers an r -type pharmacological effect can be formally defined as the scoring function:

$$f(v_i, r, v_j) = \sigma(\mathbf{z}_i \mathbf{D}_r \mathbf{R} \mathbf{D}_r \mathbf{z}_j^T), \quad (3)$$

where \mathbf{z}_i and \mathbf{z}_j are the $1 \times k$ embedding vectors of drug nodes v_i and v_j respectively, \mathbf{D}_r is a $k \times k$ feature importance diagonal matrices concerning type r , \mathbf{R} is a $k \times k$ feature association matrix across different interaction types, $\sigma(\cdot)$ is the Sigmoid function that converts the confidence score of being an r -type interaction into a probability value of $[0,1]$.

2.5 loss function

The encoder \mathcal{F}_e and the decoder \mathcal{F}_d can be trained as an end-to-end model of multi-type DDI prediction. The loss function of MTDDI is composed of two components. The first one measures the difference between the original multi-type interaction network G and the reconstructed network \tilde{G} . The second one is a regularization item, which keeps the similar drugs as close as possible in the embedding space.

Let a_{ij}^r be the true label of a triplet (v_i, r, v_j) for the pair of drug d_i and drug d_j in the r -th slice network G_r , and p_{ij}^r be the predicted probability of being an interaction of type r . For the r -th slice network G_r , its loss function L_{ij}^r is defined by a binary cross entropy as follows:

$$L_{ij}^r = -\sum_{i,j} (a_{ij}^r \log(p_{ij}^r) + (1 - a_{ij}^r)(1 - \log(p_{ij}^r))). \quad (4)$$

The positive samples are taken as the interactions in G_r while the negative samples are randomly sampled among its unlabeled drug pairs. The number of negative

samples is same as that of positive samples. For all the sliced networks, the global loss function is defined as $L = \sum_{r=1}^m L_{ij}^r$.

Let $\mathbf{S} = \{s_{i,j}\} \in [0,1], i, j = 1, 2, \dots, n$ be the drug similarity matrix. The regularization item is defined as:

$$Reg = \sum_{i,j=1}^n (s_{i,j} \cdot \|\mathbf{z}_i - \mathbf{z}_j\|_2^2), \quad (5)$$

where \mathbf{z}_i and \mathbf{z}_j are the embedding representations of drug d_i and drug d_j generated by the encoder respectively. It can be written in an elegant matrix form as follows:

$$Reg = 2\alpha \cdot \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}), \quad (6)$$

where \mathbf{Z} is an $n \times k$ feature matrix stacked by feature vectors, $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is a Laplace matrix, \mathbf{D} is an $n \times n$ diagonal matrix derived from \mathbf{S} and its element $D_{i,i} = \sum_j s_{i,j}$. This regularization item utilizes pre-defined drug similarities to constrain similar drugs as close as possible in the embedding space. This idea is similar as that in literature [35].

Therefore, the final loss of MTDDI is as follows:

$$Loss = \sum_{r=1}^m L_{ij}^r + 2\alpha \cdot \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}), \quad (7)$$

where α is a hyper parameter to adjust the weight of similarity constraint in the training phase.

2.6 Assessment

In order to measure the performance of MTDDI, the whole DDIs dataset is randomly split into a training set, a validation set and a testing set. The training set is used to train the learning model and the validation set is used to tune the model to ensure an optimal predictive performance. The testing set is used to measure the

generalization performance of the model on unlabeled data. In each experiment, we use 75% samples of the DDIs datasets as the training set, 5% samples as the validation data, and the remaining 20% samples as the testing data. The splitting process is usually repeated many times (e.g., 20 times) with different random seeds and the average performance of these repetitions is reported as the final performance.

Since our task is a multi-type prediction problem, a group of metrics is used to measure the prediction, including the area under the receiver operating characteristic curve (AUC), the area under the precision-recall curve (AUPR), Accuracy, Recall, Precision, and F1-score. Remarkably, Recall, Precision, and F1-score have their macro versions and micro versions, respectively. Macro metrics reflect the average performance across different interaction types. For example, Macro Precision is defined as the average of the Precision values of different interaction types. In contrast, Micro metrics is analogous to corresponding metrics in binary classification by summing the numbers of true positive, false positive, true negative and false negative samples across all interaction types, respectively. Their definitions are as follows:

$$\text{Accuracy} = \frac{1}{l} \sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}, \quad (8)$$

$$\text{Macro Precision} = \frac{1}{l} \sum_{i=1}^l \frac{TP_i}{TP_i + FP_i}, \quad (9)$$

$$\text{Macro Recall} = \frac{1}{l} \sum_{i=1}^l \frac{TP_i}{TP_i + FN_i}, \quad (10)$$

$$\text{Macro } F_1 = \frac{2 \times \text{macro Precision} \times \text{Macro Recall}}{\text{Macro Precision} + \text{Macro Recall}}, \quad (11)$$

$$\text{Micro Precision} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i}, \quad (12)$$

$$\text{Micro Recall} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i}, \quad (13)$$

$$Micro F_1 = \frac{2 \times micro Precision \times Micro Recall}{Micro Precision + Micro Recall}, \quad (14)$$

where TP_i , TN_i , FP_i and FN_i represent the number of true positive, true negative, false positive and false negative samples in the i -type DDI prediction, respectively; l is the number of DDI interaction types. In addition, AP@50 is employed to measure the values of Macro Precision in each DDI type on average in terms of top-50 predicted DDIs. For any of above metrics, the greater value, the better prediction.

3 Results and Discussion

We designed some experiments to address the following questions: 1) Does MTDDI improve multi-type DDI classification? 2) Can MTDDI achieve a good predictive performance in multi-type DDI prediction? 3) How both the residual strategy and the similarity regularization in the encoder help the prediction? 4) How the feature importance matrices in the decoder help finding the dependency between DDI types?

3.1. Parameter settings

To learn a good model of multi-type DDI prediction, we first determined the architecture of the encoder as follows. The one-hot encoding of 2926 nodes in the multi-type DDI network were adopted as the input features of our MTDDI. The encoder is composed of four hidden layers, in which the number of neurons is determined empirically. Besides, to accommodate the residual strategy, the second, the third and the fourth hidden layers contains the same number of neurons. Thus, the numbers of neuron in the input and four hidden layers are 2926, 1024, 128, 128 and 128, respectively.

With this encoder architecture and the tensor factorization-like decoder, we performed a grid search with Adam optimizer [36] to tune major parameters of our MTDDI, including epoch, learning rate, batch size, and the hyper parameter for the similarity regularization. The epoch, referring to as the number of training iterations, was tuned from the list of values {5, 10, 20, 30, 40, 50, 60, 70}. The learning rate,

determining whether and when the objective function converges to the optimal values, was empirically investigated from the list {0.0001, 0.001, 0.005, 0.01, 0.05, 0.1}. A mini-batch strategy, sampling a fixed number of drug pairs in each batch, was tuned from {50, 100, 200, 400, 600, 1000, 2000}. The hyper parameter α , adjusting the weight of similarity constraint, was examined from the list {0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5}. Finally, we experimentally determined a well-trained MTDDI by setting the epoch as 40, the learning rate as 0.001, the batch-size as 400, and the hyper parameter α as 0.02.

3.2 Comparison of our MTDDI with other three existing models in multi-type DDI classification

In order to answer the first question, we compared MTDDI with other three state-of-the-art multi-class classification models, including DeepDDI [11], Lee's model [22], and DDIMDL [12]. We only focus on these deep learning-based models because they have demonstrated superior performance to regular shallow models. In common, these methods first treat rows in a drug similar matrix as corresponding drug feature vectors, then set the concatenation of two feature vectors as the feature vector of a DDI, and last train a multi-layer DNN with both feature vectors and types of DDIs as the classifier. Differently, in terms of model architecture, DeepDDI is a model for homogeneous interaction feature (i.e., chemical structure) whereas both Lee's model and DDIMDL are two models for accommodating heterogeneous DDI features (e.g., pathway, GO terms and binding proteins).

Moreover, to cope with the high dimension of DDI feature, they utilized various tricks to enhance their models. DeepDDI [11] employed the Principal Component Analysis (PCA) to reduce the feature dimension before training the nine-layer DNN. Lee et al [22] first utilized three three-layer autoencoder for three source of raw DDI features respectively, and then concatenated three sources of dimension-reduced features as the training feature of the eight-layer DNN. DDIMDL [12] trained four

three-layer DNNs for four sources of DDI features respectively, and averaged the individual predictions of those trained DNNs as the final prediction.

These methods are designed for the classification of multi-type DDIs, and they determine the pharmacological effect type for a given DDIs, while our MTDDI exceeds the task with the direct discrimination of whether an unknown drug pair results in one or more pharmacological effects of interest. Thus, our MTDDI is accommodated to the version of multi-class classification task. In detail, all DDIs are divided into the training samples, and the test samples. For each type, the DDIs belonging to this DDI type are considered as the positive samples, and the DDIs not belonging to this DDI type are considered as the negative samples. We implemented DeepDDI, Lee’s model and DDIMDL with their published source codes and the default parameters.

The comparison results in Table 1 show that our MTDDI achieves the best performance with the significant improvements of 0.2~1.2%, 4.7~8.6%, 2.9~11.7%, 1.2~11.7%, 0.3~1.5%, 1.2~11.7% against other state-of-the-art methods in terms of Macro AUC, Macro AUPR, Micro Recall, Micro Precision, Accuracy, and Micro F1-score respectively. Even in terms of Micro AUC, it still achieves a comparable result (0.980), only 0.007 lower than the best baseline. The results demonstrate the superiority of MTDDI.

Table1. Comparison results of MTDDI with three state-of-the-art models in multi-type DDI classification

Model	Micro AUC	Macro AUC	Macro AUPR	Micro Recall	Micro Precision	accuracy	micro F ₁
DeepDDI	0.985	0.972	0.733	0.685	0.788	0.961	0.788
Lee’s	0.987	0.982	0.694	0.683	0.683	0.949	0.683
DDIMDL	0.983	0.975	0.708	0.771	0.771	0.958	0.771
MTDDI	0.980	0.984	0.780	0.800	0.800	0.964	0.800

3.3 Performance of MTDDI in multi-type DDI prediction

Existing methods only consider the multi-class classification of single-fold interactions, where a DDI triggers only one pharmacological effect. However, ~20% DDIs are of multiple-fold interactions, where a DDI causes two or more related pharmacological effects (see also Section 2.1). Since the classification of multiple-fold interactions requires multi-label classifiers, existing methods cannot handle this task. Owing to the decoder, our MTDDI is capable to address the issue of predicting the multi-fold interactions. In this sense, we run a multi-type DDI prediction to demonstrate the good predictive performance of MTDDI. The results in Table2 show that our MTDDI can effectively predict the single-fold and multi-fold DDIs.

Table2. Performance of MTDDI for multi-type DDIs prediction in both single-fold and multi-fold DDIs

	Relation types	AUC	AUPR	Accuracy	Precision	Recall	F1	AP@50
	Absorption	0.989	0.982	0.967	0.954	0.982	0.968	0.952
	Metabolism	0.948	0.932	0.898	0.850	0.943	0.894	0.920
	Serum Concentration	0.965	0.938	0.937	0.920	0.971	0.939	0.956
	Excretion	0.937	0.860	0.904	0.867	0.971	0.940	0.820
	Synergy Activity	0.995	0.992	0.980	0.975	0.985	0.980	0.987
single-fold	Antagonism Activity	0.986	0.979	0.959	0.944	0.977	0.960	0.974
	Toxicity Activity	0.970	0.968	0.923	0.929	0.931	0.923	0.985
	Adverse Effect	0.980	0.966	0.957	0.937	0.980	0.958	0.907
	Antagonism Effect	0.992	0.989	0.972	0.963	0.982	0.978	0.971
	Synergy Effect	0.995	0.993	0.9778	0.970	0.985	0.978	0.995
	PD triggered by PK	0.990	0.986	0.968	0.954	0.984	0.969	0.989
multi-fold	Macro metrics	0.977	0.962	0.949	0.933	0.972	0.953	0.951

In order to further verify the performance of our MTDDI to predict the new DDIs and their interaction types in unknown DDIs, The inspiring prediction impels us to

perform a novel transductive inference of potential DDIs among all drug pairs and their interaction types. Such an inference validates the performance of MTDDI in practice. To accomplish this task, we first used the whole dataset with known DDIs to train MTDDI, then employed the trained MTDDI to infer how likely unlabeled drug pairs trigger specific pharmacological effects among 11 interaction types. After that, we ranked these unlabeled drug pairs in each interaction type according to their type-specific predicting scores. Finally, we picked up top-20 type-specific candidates in each interaction type and validated them by both the latest version of DrugBank (version 5.1.8, on January 18, 2021) and the online Drug Interaction Checker tool (Drugs.com).

The validation was performed in both single-fold interactions and multi-fold interactions respectively. In the prediction results of single-fold interactions, 40 out of 220 predicted DDI candidates (18.2%) are confirmed. The average rank of 40 verified DDIs is 7.75, indicating that our MTDDI can effectively detect the potential DDIs as well as different types of DDIs. The detailed results can be found in Table S1 of supplement file. We further picked up some validated DDI candidates (i.e., Case 31, Case 34, Case15 and Case 16) to show how DDI prediction contributes to synergistic drug combination and drug contraindication. For example, when two drugs of Pregabalin and Benmoxin are combined, the therapeutic efficacy of Benmoxin can be increased (Case 31). In addition, the therapeutic efficacy of Mebanazine can be increased when used in combination with Pregabalin (Case 34). In contrast, the risk or severity of QTc prolongation can be increased when Quinidine is combined with Promethazine (Case 15). Besides, the risk or severity of serotonin syndrome can be increased when Linezolid is combined with Ergotamine (Case 16). These results manifest that the MTDDI can provide a preliminary screening for synergistic drug combination and drug contraindication.

In the prediction results of multi-fold interactions, 17 out of 50 two-fold predicted candidates and 8 out of 60 three-fold predicted candidates are confirmed, respectively. The detailed results are listed in Table S2 of supplement file. As illustrated, we picked

up a two-fold interaction case (Case 8) and a three-fold interaction case (Case 18) to show how MTDDI contributes to find multi-fold interaction cases. For the example of two-fold interaction, DrugBank states “Acebutolol may increase the arrhythmogenic activities of Digoxin.” , while DDI Checker states “Using Acebutolol together with Digoxin may slow your heart rate and lead to increased side effects.” (Case 8). Both statements show that the pair of Digoxin and Acebutolol triggers a PK antagonistic activity and further results in a PD adverse effect. For the example of three-fold interaction, two statements are similarly found, but contain three pharmacological effects as follows “Voriconazole may increase the blood levels and effects of Trazodone” and “The risk or severity of QTc prolongation can be increased when Trazodone is combined with Voriconazole” (Case 18). The pair of Voriconazole and Trazodone increases both PK serum and PD synergy of Trazodone, but also increases the risk of adverse effects as well. In total, these newly-predicted multi-type DDIs demonstrate the potentials of our MTDDI in practice.

3.4 Influence of hidden layers, residual strategy and similarity regularization in encoder

In this section, we investigated how three factors (i.e., the number of hidden layers, the similarity regularization, and the residual strategy) in the encoder affect the performance of MTDDI. First, after removing the similarity regularization and the residual strategy in MTDDI, we adopted MTDDI two variants, that is, MTDDI with 2 hidden layers (denoted as MTDDI-2) and 4 hidden layers (denoted as MTDDI-4). The two hidden layers in MTDDI-2 contain [1024, 128] neurons, and four hidden layers in MTDDI-4 contain [1024, 128, 128, 128] neurons, respectively. On the architecture of MTDDI-4, we added the residual strategy to generate an additional variant of MTDDI (denoted as MTDDI-4-R). If the similarity regularization is further added to MTDDI-4-R, the variant of MTDDI-4-R is the full architecture of MTDDI. The results of MTDDI and its variants (i.e., MTDDI-2, MTDDI-4 and MTDDI-4-R) are shown in Table 3, from which we can obtain three following crucial points.

- (1) MTDDI-4 is worse than MTDDI-2 in all the measuring metrics. Obviously,

the increment of the number of hidden layers decreases the predictive performance because of the “over smoothing” issue derived from GCN.

- (2) Compared with MTDDI-2 and MTDDI-4, MTDDI-4-R owing to the residual strategy achieves the significant improvement. Thus, the residual strategy can relax the “over smoothing” issue in the case of deeper GCN architecture.
- (3) Compared with these variants, the full architecture of MTDDI having the additional similarity regularization further improves the prediction. Thus, the similarity regularization helps constrain similar drugs as close as possible in the embedding space to cope with the issue that missing interaction label between similar drugs causes their remoteness in the network.

In summary, with the help of residual strategy, MTDDI can accommodate deep GCN architecture (e.g., containing >2 layers). Also, its similarity regularization further helps capture missing interactions.

Table3. Performance of similarity constraint and residual strategy in MTDDI

Variant	Macro	Macro	accuracy	Macro	Macro	Macro	AP@50
	AUC	AUPR		Precision	Recall	F1	
MTDDI-2	0.953	0.913	0.906	0.881	0.958	0.916	0.909
MTDDI-4	0.928	0.901	0.878	0.847	0.925	0.885	0.838
MTDDI-4-R	0.971	0.954	0.948	0.927	0.973	0.949	0.959
MTDDI	0.977	0.962	0.949	0.933	0.972	0.953	0.951

3.5 Influence of different implementations of decoder

Since the decoder in MTDDI is loosely coupled with the encoder, we should adopt various decoder models. In this section, we compared three implementations of the decoder, including the inner production $\mathbf{z}_i \mathbf{z}_j^T$ in the traditional GCN, the

type-specific association $\mathbf{z}_i \mathbf{M}_r \mathbf{z}_j^T$ in R-GCN, as well as our type-specific importance association $\mathbf{z}_i \mathbf{D}_r \mathbf{R} \mathbf{D}_r \mathbf{z}_j^T$. According to their original algorithms, these three implementations are denoted as InnerProd, RESCAL and DEDICOM, respectively. See Section 2.3.2 Decoder for details.

The comparison results in Table 4 show that InnerProd is the worst and DEDICOM is the best.. The potential reason of DEDICOM significantly outperforming two other models is as follows. The inner production $\mathbf{z}_i \mathbf{z}_j^T$ only indicates how likely drug d_i interacts with drug d_j , but it cannot model interaction types. In contrast, RESCAL reflects the difference between interaction types and models the likelihood of being a type-specific interaction by m additional type-specific feature association matrices $\{\mathbf{M}_r\}$. Compared with RECAL, to indicate how likely the pair of drug d_i and drug d_j triggers an r -type pharmacological effect, DEDICOM employs a global feature association matrix \mathbf{R} , as well as m additional type-specific diagonal matrices $\{\mathbf{D}_r\}$, which reflects that feature importance varies across interaction types.

Table4. Performance of different implementations in decoder

Model	AUC	Macro AUPR	Accuracy	Macro Precision	Macro Recall	Macro F1	AP@50
InnerProd	0.634	0.588	0.591	0.571	0.947	0.704	0.125
RESCAL	0.896	0.879	0.832	0.797	0.917	0.850	0.869
DEDICOM	0.977	0.962	0.949	0.933	0.972	0.953	0.951

To further validate whether feature importance matrices capture the dependency between DDI types, we calculated the pairwise correlations among matrices $\{\mathbf{D}_r\}$. First, we calculated their correlations by diagonal vectors of matrices $\{\mathbf{D}_r\}$ since these matrices are diagonal matrices (Figure 3). Then, we categorized these types into a

pharmacokinetic group (PK) and a pharmacodynamic group (PD) in terms of their pharmacological behaviors. The PK group contains the first 7 types while the remaining types belong to the PD group. After that, we calculated the average values of absolute correlations within PK and within PD (denoted as C_{PK} and C_{PD}), and the average value of absolute correlations between PK and PD (denoted as C_B), respectively. The results reveal that C_{PK} (0.264) is significantly greater than C_{PD} (0.086), and C_B (0.344) is the greatest. Similarly, we calculated the average values (in Figure 3) of absolute correlations within the individual DDI types, and found that Absorption is the maximum (0.301) and Antagonism Effect is the minimum (0.074). Moreover, we enumerated the correlations between individual DDI types. For example, Absorption is significantly related with Serum Concentration ($\rho = -0.55$) and Toxicity Activity ($\rho = -0.45$), respectively; Synergy Activity is significantly related with Synergy Effect ($\rho = -0.53$); Antagonism Effect is independent to Synergy Effect ($\rho = -0.005$). All the p-values of correlation entries are significantly less than 0.0001. Totally, the results in Figure 3 demonstrate that DDI types are not independent to each other, and some of them show significant correlations. Thus, the feature importance matrices in the decoder can capture the dependency relation of DDI types in some sense, and they would contribute uncovering the forming mechanism of DDI as well as finding potential synergistic drug combinations with the aid of more medical knowledges.



Figure 3. Heat map of correlation analysis for different DDI types.

4 Conclusions

This work proposed a novel end-to-end deep learning-based model (called MTDDI) for predicting DDIs as well as its types and exploring the underlying mechanism of DDIs. MTDDI designs an encoder by the enhanced deep relational graph convolutional networks to capture the structural relationship between multiple-type DDI entries. Meanwhile, MTDDI also employs a tensor-like decoder to uniformly model both single-fold interactions and multi-fold interactions, such that it can directly discriminate whether an unlabeled type-specific drug pair results in one or more pharmacological effects. The superiority of MTDDI is demonstrated by

comparing with state-of-the-art deep learning-based models in the task of DDI classification. Furthermore, MTDDI achieves an inspiring performance for predicting multi-type DDIs in the case of both single-fold DDIs and multi-fold DDIs, and its practical capability of predicting multi-type DDIs is further validated in case study of both single-fold DDIs and multi-fold DDIs by the latest version of DrugBank and an online Drug Interaction Checker tool. More importantly, MTDDI utilizes a set of type-specific feature importance matrices to reveal the dependency between DDI types by calculating their correlations. This crucial observation gives an insight to uncover the mechanism of DDIs, which are beneficial for the inference of drug combination in treating complex diseases.

Abbreviations

DDIs: Drug-Drug Interactions; GCN: Graph Convolution Network; DBPs: Drug-Binding Proteins; AUC: Area Under the receiver operating characteristic Curve; AUPR: Area Under the Precision-Recall curve; ACC: ACCuracy; R-GCN: Relational Graph Convolution Network; SMILES: Simplified Molecular Input Line Entry System; MACCSkeys: Molecular ACCess System keys; PK: PharmacoKinetic; PD: PharmacoDynamic

Declarations

Ethics approval and consent to participate

No ethics approval was required for the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study and the code of MTDDI are openly available at the website of <https://github.com/NWPU-903PR/MTDDI>.

Competing interests

None of the authors has any competing interests.

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). The funding body did not play any roles in the design of the study, collection, analysis, and interpretation of data, and in writing the manuscript.

Authors' contributions

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

Acknowledgments

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

References

1. Cheng F KI, Barabási AL.: **Network-based prediction of drug combinations.** . *Nat Commun* 2019, **10(1):1197**.
2. Niu J SR, Mager DE. : **Pharmacodynamic Drug-Drug Interactions.** . *Clin Pharmacol Ther* 2019, **105(6)** :1395-1406.
3. 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**.
4. Takeda T, Hao M, Cheng T, Bryant SH, Wang Y: **Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge.** *J Cheminform* 2017, **9:16**.
5. Sridhar D, Fakhraei S, Getoor: **A Probabilistic Approach for Collective Similarity-based Drug-Drug Interaction Prediction.** *Bioinformatics* 2016, **32(20):3175-3182**.
6. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X: **Predicting potential drug-drug interactions by**

- integrating chemical, biological, phenotypic and network data.** *BMC bioinformatics* 2017, **18**(1):18.
7. Wen. Z, Kanghong. J, Feng. H, Yanlin. C, Bolin. L, Jinghao. L, . GJ: **SFLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions.** *J Information Sciences* 2019, **497**:189-201.
 8. Andrej K, Polonca F, Brane Le, Jinn-Moon Y: **Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning.** *Plos One* 2018, **13**(5):e0196865-.
 9. Yu H, Mao KT, Shi JY, Huang H, Chen Z, Dong K, Yiu SM: **Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization.** *BMC Syst Biol* 2018, **12**(Suppl 1):14.
 10. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R: **INDI: a computational framework for inferring drug interactions and their associated recommendations.** *Mol Syst Biol* 2012, **8**(1).
 11. Ryu JY, Kim HU, Lee SY: **Deep learning improves prediction of drug-drug and drug-food interactions.** *Proc Natl Acad Sci U S A* 2018, **115**(18):E4304-E4311.
 12. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S: **A multimodal deep learning framework for predicting drug-drug interaction events.** *Bioinformatics* 2020, **36**(15):4316-4322.
 13. Cheng F, Zhao Z: **Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties.** *J Am Med Inform Assoc* 2014, **21**(e2):e278-286.
 14. 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.
 15. Shtar G, Rokach L, Shapira B: **Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures.** *PLoS One* 2019, **14**(8):e0219796.
 16. Shi J-Y, Mao K-T, Yu H, Yiu S-M: **Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization.** *Journal of cheminformatics* 2019, **11**(1).
 17. Ferdousi R, Safdari R, Omid Y: **Computational prediction of drug-drug interactions based on drugs functional similarities.** 2017, **70**:54.
 18. Rohani N, Eslahchi C, Katanforoush A: **ISCMF: Integrated similarity-constrained matrix factorization for drug–drug interaction prediction.** *Network Modeling Analysis in Health Informatics and Bioinformatics* 2020, **9**(1).
 19. Jing Y, Bian Y, Hu Z, Wang L, Xie X: **Deep Learning for Drug Design: an Artificial Intelligence Paradigm for Drug Discovery in the Big Data Era.** *AAPS J* 2018, **20**(4):79.

20. 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, **179**:37-46.
21. Deepika SS, Geetha TVJJoBI: **A meta-learning framework using representation learning to predict drug-drug interaction.** *J Biomed Inform* 2018, **84**(84):136-147.
22. Lee G PC, Ahn J.: **Novel deep learning model for more accurate prediction of drug-drug interaction effects.** *BMC Bioinformatics* 2019 **20**:415.
23. 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.
24. Jonker DM VS, van der Graaf PH, Voskuyl RA, Danhof M. : **Towards a mechanism-based analysis of pharmacodynamic drug-drug interactions in vivo.** *Pharmacol Ther* 2005, **106**(1):1-18.
25. Seung H, Lee, et al.: **Pharmacokinetic and pharmacodynamic insights from microfluidic intestine-on-a-chip models.** *Expert opinion on drug metabolism & toxicology* 2019, **15**(12):1005-1019.
26. Rogers D HM: **Extended-connectivity fingerprints.** *J Chem Inf Model* 2010, **50** (5) :742-754.
27. Zitnik M, Agrawal M, Leskovec J: **Modeling polypharmacy side effects with graph convolutional networks.** *Bioinformatics* 2018, **34**(13):i457-i466.
28. Schlichtkrull M KTN, Bloem P , et al: **Modeling relational data with graph convolutional networks.** *Springer, Cham* 2017.
29. Kipf TN, Welling M: **Semi-supervised classification with graph convolutional networks.** 2016, **arXiv:1609.02907**.
30. Kipf TN, Welling M: **Variational Graph Auto-Encoders.** 2016, **arXiv:1611.07308**.
31. Nickel M, Tresp V, Kriegel HP: **A Three-Way Model for Collective Learning on Multi-Relational Data.** In: *International Conference on International Conference on Machine Learning: 2011*.
32. EE Papalexakis CF, N. D. Sidiropoulos: **Tensors for Data Mining and Data Fusion: Models, Applications, and Scalable Algorithms.** *ACM Trans Intell Syst Technol* 2016, **8**(2157-6904):44.
33. Yue X WZ, Huang J , et al.: **Graph Embedding on Biomedical Networks: Methods, Applications, and Evaluations.** *Bioinformatics* 2019, **36**(Suppl 1).
34. Yu Z HF, Zhao X, Xiao W, Zhang W. : **Predicting drug-disease associations through layer attention graph convolutional network.** *Brief Bioinform* 2020, **243**.
35. Wang D, Peng C, Zhu W: **Structural Deep Network Embedding.** In: *Acm Sigkdd International*

Conference on Knowledge Discovery & Data Mining: 2016.

36. Kingma DaB, J.: **Adam: A method for stochastic optimization**. 2014, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).

Table S1. Investigation of novel single-fold DDIs predicted by MTDDI

	Drug1 ID	Drug2 ID	Drug1 name	Drug2 name	DDI type	Rank	Description in latest Drugbank
1	DB01098	DB01390	Rosuvastatin	Sodium bicarbonate	PK absorption	3	Sodium bicarbonate can cause a decrease in the absorption of Rosuvastatin resulting in a reduced serum concentration and potentially a decrease in efficacy.
2	DB00275	DB08957	Olmesartan	Hexoprenaline	PK activity antagonist	1	Hexoprenaline may decrease the antihypertensive activities of Olmesartan.
3	DB00280	DB04836	Disopyramide	Amineptine	PK activity antagonist	4	Amineptine may decrease the hypoglycemic activities of Disopyramide.
4	DB00708	DB00966	Sufentanil	Telmisartan	PK activity antagonist	7	Sufentanil may decrease the antihypertensive activities of Telmisartan.
5	DB00876	DB08957	Eprosartan	Hexoprenaline	PK activity antagonist	12	Hexoprenaline may decrease the antihypertensive activities of Eprosartan.
6	DB01442	DB15120	MMDA	GSK-239512	PK activity antagonist	18	MMDA may decrease the sedative and stimulatory activities of GSK-239512.
7	DB05271	DB09194	Rotigotine	Etoferidone	PK activity synergy	4	Etoferidone may increase the sedative activities of Rotigotine.
8	DB01624	DB06762	Zuclopenthixol	Pinacidil	PK activity synergy	5	Zuclopenthixol may decrease the antihypertensive activities of Pinacidil.
9	DB01203	DB06762	Nadolol	Pinacidil	PK activity synergy	8	Nadolol may increase the hypotensive activities of Pinacidil.
10	DB01203	DB13620	Nadolol	Potassium gluconate	PK activity synergy	10	Potassium gluconate may increase the hyperkalemic activities of Nadolol.

11	DB01119	DB09242	Diazoxide	Moxonidine	PK activity synergy	14	Diazoxide may increase the hypotensive activities of Moxonidine.
12	DB04918	DB09285	Ceftobiprole	Morniflumate	PK activity toxicity	2	The risk or severity of nephrotoxicity can be increased when Ceftobiprole is combined with Morniflumate.
13	DB00372	DB00898	Thiethylperazine	Ethanol	PK activity toxicity	5	Thiethylperazine may increase the central nervous system depressant (CNS depressant) activities of Ethanol.
14	DB00372	DB00679	Thiethylperazine	Thioridazine	PK activity toxicity	8	The risk or severity of adverse effects can be increased when Thiethylperazine is combined with Thioridazine.
15	DB00908	DB01069	Quinidine	Promethazine	PD adverse effect	1	The risk or severity of QTc prolongation can be increased when Quinidine is combined with Promethazine.
16	DB00601	DB00696	Linezolid	Ergotamine	PD adverse effect	3	The risk or severity of serotonin syndrome can be increased when Linezolid is combined with Ergotamine.
17	DB00584	DB01086	Enalapril	Benzocaine	PD adverse effect	7	Fluvoxamine may decrease the excretion rate of Gadoteric acid which could result in a higher serum level.
18	DB00311	DB01197	Ethoxzolamide	Captopril	PD adverse effect	9	The excretion of Captopril can be decreased when combined with Ethoxzolamide.
19	DB06288	DB12086	Amisulpride	Oxitropium	PD adverse effect	10	Amisulpride may increase the central nervous system depressant (CNS depressant) activities of Oxitropium.
20	DB00190	DB01239	Carbidopa	Chlorprothixene	PK excretion	5	The therapeutic efficacy of Carbidopa can be decreased when used in combination with Chlorprothixene.
21	DB00214	DB00435	Torasemide	Nitric Oxide	PK excretion	11	Nitric Oxide may increase the excretion rate of Torasemide which could result in a lower serum level and potentially a reduction in efficacy.
23	DB00201	DB14010	Caffeine	5-methoxy-N,N-dimethyltryptamine	PK metabolism	5	The metabolism of Perphenazine can be decreased when combined with Felodipine.
24	DB01016	DB01388	Glyburide	Mibefradil	PK metabolism	8	The metabolism of Glyburide can be decreased when combined with

							Mibefradil.
25	DB00338	DB01110	Omeprazole	Miconazole	PK metabolism	17	The metabolism of Omeprazole can be decreased when combined with Miconazole.
26	DB00599	DB01041	Thiopental	Thalidomide	PD antagonist	3	Thiopental may increase the central nervous system depressant (CNS depressant) activities of Thalidomide.
27	DB00925	DB06701	Phenoxybenzamine	Dexmethylphenidate	PD antagonist	7	Dexmethylphenidate may decrease the antihypertensive activities of Phenoxybenzamine.
28	DB00725	DB09268	Homatropine methylbromide	Picosulfuric acid	PD antagonist	10	The therapeutic efficacy of Picosulfuric acid can be decreased when used in combination with Homatropine methylbromide.
29	DB00280	DB06150	Disopyramide	Sulfadimethoxine	PD synergy	3	The therapeutic efficacy of Disopyramide can be increased when used in combination with Sulfadimethoxine.
30	DB00653	DB12131	Magnesium sulfate	Vinpocetine	PD synergy	5	The therapeutic efficacy of Vinpocetine can be increased when used in combination with Magnesium sulfate.
31	DB00230	DB09246	Pregabalin	Benmoxin	PD synergy	9	The therapeutic efficacy of Benmoxin can be increased when used in combination with Pregabalin.
32	DB00230	DB09248	Pregabalin	Mebanazine	PD synergy	10	The therapeutic efficacy of Mebanazine can be increased when used in combination with Pregabalin.
33	DB00230	DB09249	Pregabalin	Octamoxin	PD synergy	14	The therapeutic efficacy of Octamoxin can be increased when used in combination with Pregabalin.
34	DB00232	DB01359	Methyclothiazide	Penbutolol	PD synergy	18	The therapeutic efficacy of Penbutolol can be increased when used in combination with Methyclothiazide.
35	DB01016	DB01582	Glyburide	Sulfamethazine	PD synergy	19	The therapeutic efficacy of Glyburide can be increased when used in combination with Sulfamethazine.
36	DB00254	DB01394	Doxycycline	Colchicine	PK and PD	1	Doxycycline may decrease the excretion rate of Colchicine which could result in a higher serum level.
37	DB00182	DB00798	Amphetamine	Gentamicin	PK and PD	6	Gentamicin may decrease the excretion rate of Amphetamine which could

							result in a higher serum level.
38	DB00091	DB12278	Cyclosporine	Propiverine	PK serum	6	Cyclosporine may decrease the excretion rate of Propiverine which could result in a higher serum level.
39	DB00631	DB01142	Clofarabine	Doxepin	PK serum	9	Clofarabine may decrease the excretion rate of Doxepin which could result in a higher serum level.
40	DB00682	DB08864	Warfarin	Rilpivirine	PK serum	13	The serum concentration of Warfarin can be increased when it is combined with Rilpivirine.

Table S2. Investigation of novel multi-fold DDIs predicted by MTDDI

	Drug1 ID	Drug2 ID	Drug1 name	Drug2 name	Predicted DDIs type	Benchmark	Description in Benchmark
1	DB00218	DB00322	Moxifloxacin	Floxuridine	PK absorption, PK serum	drugs.com	Limited data suggest that chemotherapy with antineoplastic agents may reduce the plasma concentrations of oral quinolone antibiotics. The proposed mechanism is decreased quinolone absorption secondary to alteration of intestinal mucosa by cancer chemotherapy.
2	DB00219	DB13136	Moxifloxacin	Fluindione	PD synergistic effect, PK absorption	DrugBank	The therapeutic efficacy of Fluindione can be increased when used in combination with Moxifloxacin.
3	DB00448	DB01414	Lansoprazole	Cefacetile	PK excretion, PD adverse effect, PKD	DrugBank	The excretion of Cefacetile can be decreased when combined with Lansoprazole.
4	DB00468	DB01081	Quinine	Diphenoxylate	PD adverse effect, PK toxic activity	DrugBank	The risk or severity of adverse effects can be increased when Quinine is combined with Diphenoxylate.
5	DB00608	DB0651	Chloroquine	Dyphylline	PK excretion, PK serum, PD adverse effect	DrugBank	Chloroquine may decrease the excretion rate of Dyphylline which could result in a higher serum level.
6	DB06724	DB14488	Calcium carbonate	Ferrous gluconate	PK serum, PD	DrugBank	Using calcium carbonate together with ferrous gluconate may decrease the effects of ferrous gluconate.

					synergistic effect		
7	DB00230	DB00535	Pregabalin	Cefdinir	PK excretion, PK serum, PD adverse effect	DrugBank	Pregabalin may decrease the excretion rate of Cefdinir which could result in a higher serum level.
8	DB00390	DB01193	Digoxin	Acebutolol	PD adverse effect, PK toxic activity	DrugBank drugs.com	Acebutolol may increase the arrhythmogenic activities of Digoxin. (drugbank) Using acebutolol together with digoxin may slow your heart rate and lead to increased side effects. (drugs.com)
9	DB00395	DB0181	Carisoprodol	Diphenoxylate	PD adverse effect, PKD	DrugBank	The risk or severity of adverse effects can be increased when Carisoprodol is combined with Diphenoxylate.
10	DB00422	DB06701	Methylphenidate	Dexmethylphenidate	PD adverse effect, PKD	DrugBank	The risk or severity of hypertension can be increased when Methylphenidate is combined with Dexmethylphenidate.
11	DB00424	DB01081	Hyoscyamine	Diphenoxylate	PD adverse effect, PKD	DrugBank	The risk or severity of adverse effects can be increased when Diphenoxylate is combined with Hyoscyamine.
12	DB00448	DB01414	Lansoprazole	Cefacetriple	PK excretion, PK serum, PD adverse effect	DrugBank	The excretion of Cefacetriple can be decreased when combined with Lansoprazole.
13	DB00468	DB01081	Quinine	Diphenoxylate	PK antagonistic activity, PD adverse	DrugBank	The risk or severity of adverse effects can be increased when Quinine is combined with Diphenoxylate.

					effect		
14	DB00501	DB00651	Cimetidine	Dyphylline	PK metabolism, PK serum	DrugBank	The metabolism of Dyphylline can be decreased when combined with Cimetidine.
15	DB00679	DB00861	Thioridazine	Diflunisal	PD adverse effect, PKD	DrugBank	The risk or severity of hypertension can be increased when Thioridazine is combined with Diflunisal.
16	DB00679	DB01081	Thioridazine	Diphenoxylate	PD adverse effect, PKD	DrugBank	The risk or severity of hypotension and CNS depression can be increased when Thioridazine is combined with Diphenoxylate.
17	DB06724	DB14488	Calcium carbonate	Ferrous gluconate	PK antagonistic activity, PD antagonist	DrugBank drugs.com	Using calcium carbonate together with ferrous gluconate may decrease the effects of ferrous gluconate.
18	DB00582	DB00656	Voriconazole	Trazodone	PK serum, PD synergy, PD adverse effect	DrugBank drugs.com	The risk or severity of QTc prolongation can be increased when Trazodone is combined with Voriconazole. (DrugBank) Voriconazole may increase the blood levels and effects of trazodone. (drugs.com)
19	DB00708	DB00966	Sufentanil	Telmisartan	PK antagonistic activity, PD adverse effect, PKD	DrugBank	Telmisartan and Sufentanil may have additive effects in lowering your blood pressure. You may experience headache, dizziness, lightheadedness, fainting, and/or changes in pulse or heart rate.
20	DB00968	DB01037	Methyldopa	Selegiline	PK excretion, PD adverse effect	DrugBank	Combining these medications may increase your blood pressure or cause central nervous system stimulation that can result in agitation and hallucinations.
21	DB00988	DB14010	Dopamine	5-methoxy-N,N	PD adverse	DrugBank	The risk or severity of hypertension can be increased when

				-dimethyltryptamine	effect, PK excretion , PKD		Dopamine is combined with 5-methoxy-N, N-dimethyltryptamine.
22	DB01199	DB01373	Tubocurarine	Calcium	PK antagonistic activity, PD antagonist	DrugBank	The therapeutic efficacy of Calcium polycarbophil can be decreased when used in combination with Tubocurarine.
23	DB00364	DB01233	Sucralfate	Metoclopramide	PK excretion, PK serum, PD adverse effect	DrugBank	Sucralfate may decrease the excretion rate of Metoclopramide which could result in a higher serum level.
24	DB01233	DB01303	Metoclopramide	Oxtriphylline	PK metabolism, PD synergy	DrugBank	The metabolism of Oxtriphylline can be decreased when combined with Metoclopramide.
25	DB01016	DB01095	Glyburide	Fluvastatin	PK metabolism, PD synergy	DrugBank	The metabolism of Fluvastatin can be decreased when combined with Glyburide.

Figures

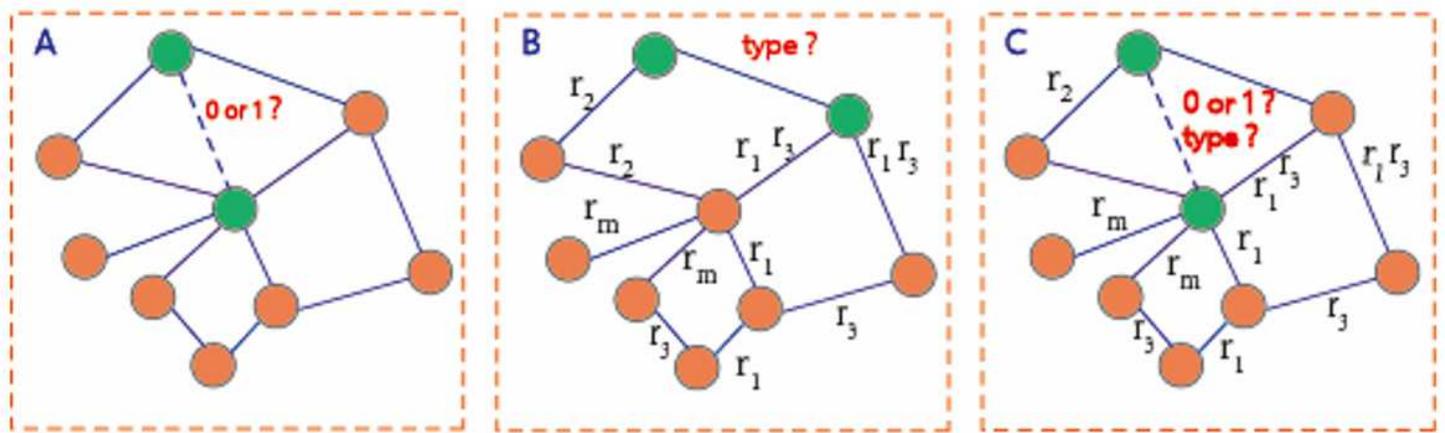


Figure 1

Three tasks in DDIs prediction: (A) Binary DDI prediction. (b) Multi-type DDI classification. (C) Multi-type DDI prediction.

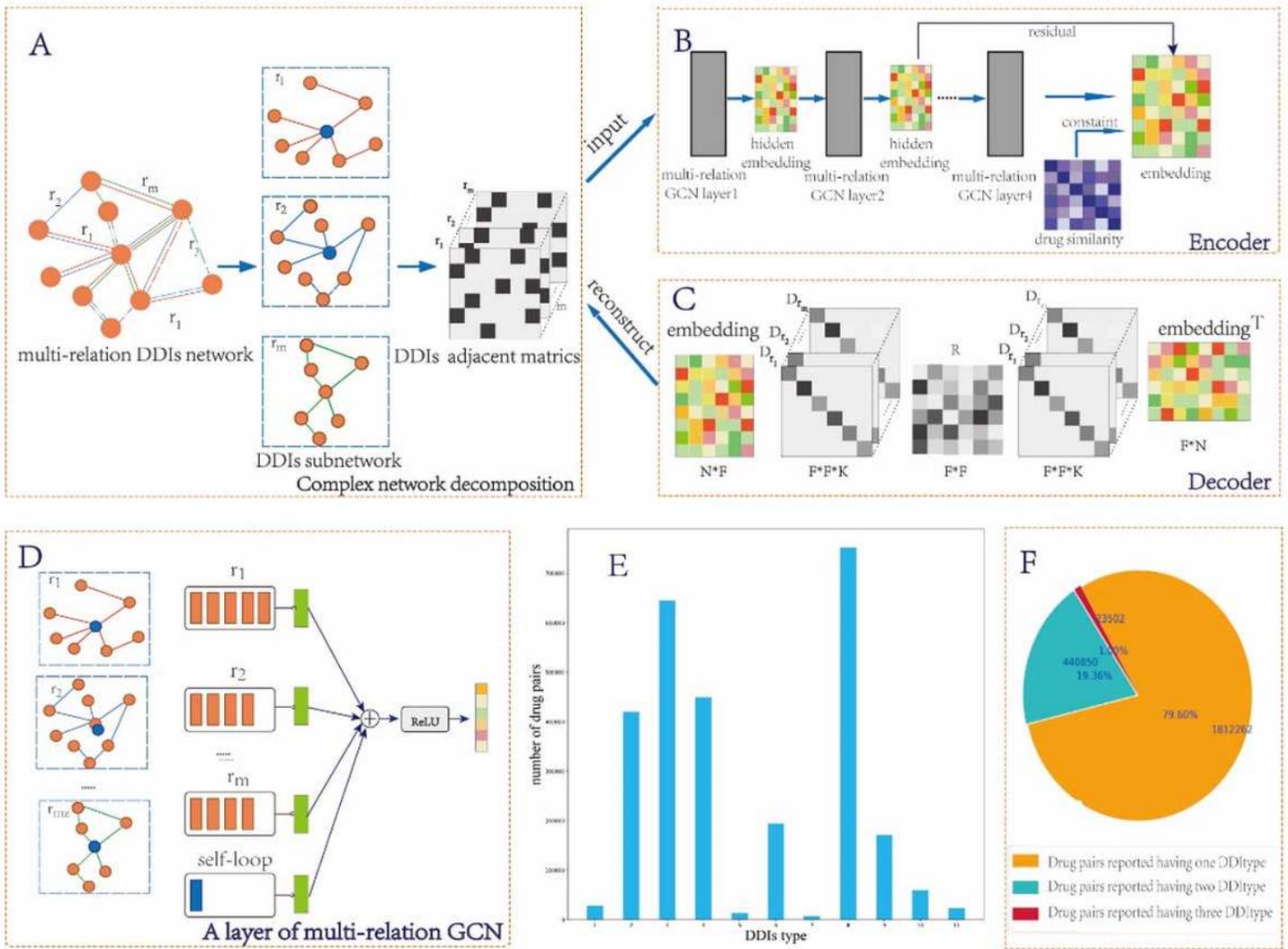


Figure 2

Overall framework of MTDDI and multi-type DDI statistics. See manuscript .pdf for full caption



Figure 3

Heat map of correlation analysis for different DDI types.