OXFORD

Briefings in Bioinformatics, 2023, **24(6)**, 1–11
https://doi.org/10.1093/bib/bbad385
Problem Solving Protocol

HetDDI: a pre-trained heterogeneous graph neural network model for drug–drug interaction prediction

Zhe Li†, Xinyi Tu†, Yuping Chen and Wenbin Lin

Corresponding authors: Wenbin Lin, School of Mathematics and Physics, University of South China, Hengyang, 421001, China. E-mail: lwb@usc.edu.cn; Yuping Chen, School of Pharmacy, University of South China, Hengyang, 421001, China. E-mail: yupingc@usc.edu.cn

†Zhe Li and Xinyi Tu contributed equally to this work.

Abstract

The simultaneous use of two or more drugs due to multi-disease comorbidity continues to increase, which may cause adverse reactions between drugs that seriously threaten public health. Therefore, the prediction of drug–drug interaction (DDI) has become a hot topic not only in clinics but also in bioinformatics. In this study, we propose a novel pre-trained heterogeneous graph neural network (HGNN) model named HetDDI, which aggregates the structural information in drug molecule graphs and rich semantic information in biomedical knowledge graph to predict DDIs. In HetDDI, we first initialize the parameters of the model with different pre-training methods. Then we apply the pre-trained HGNN to learn the feature representation of drugs from multi-source heterogeneous information, which can more effectively utilize drugs' internal structure and abundant external biomedical knowledge, thus leading to better DDI prediction. We evaluate our model on three DDI prediction tasks (binary-class, multi-class and multi-label) with three datasets and further assess its performance on three scenarios (S1, S2 and S3). The results show that the accuracy of HetDDI can achieve 98.82% in the binary-class task, 98.13% in the multi-class task and 96.66% in the multi-label one on S1, which outperforms the state-of-the-art methods by at least 2%. On S2 and S3, our method also achieves exciting performance. Furthermore, the case studies confirm that our model performs well in predicting unknown DDIs. Source codes are available at https://github.com/LinsLab/HetDDI.

Keywords: drug–drug interaction; pre-taining; heterogeneous graph neural network; multi-source information

INTRODUCTION

Human diseases often result from multifactorial etiologies, and patients may require combination therapy with two or more drugs. However, interactions between drugs can lead to adverse events, which seriously threaten the health of the public [1]. As several case reports have shown that cannabis may inhibit the metabolism of warfarin due to CYP2C9 interactions, resulting in an increase of plasma concentrations and bleeding risk [2]. Traditional channels such as drug experiments and doctors' clinical experience can obtain accurate drug–drug interactions (DDIs) but are expensive and time-consuming [3, 4]. Therefore, predicting potential DDIs quickly and effectively has become an urgent problem in clinical medicine [5].

In recent years, there has been significant progress with exciting results from the computational methods [6–10] utilized widely for predicting DDIs. These methods can be roughly classified into five categories: text mining or literature-based methods, similarity-based, feature-based, graph-based and other hybrid methods.

Text mining or literature-based methods [11–14] refer to applying natural language processing (NLP) and machine learning techniques to extract information related to DDI from extensive text data, such as scientific publications and biomedical corpus. Hong *et al.* [15] introduce a machine learning framework that enables automated biomedical relation extraction from literature repositories. Huang *et al.* [9] extract and consolidate DDIs from large-scale medical textual data. These methods have high accuracy but can only identify marked DDIs [16]. Moreover, it has faced a huge challenge as biomedical databases are updated continuously.

The similarity-based methods [17–20] are guided by the empirical assumption that drugs with similar features are more likely to have similar interactions. Ferdousi *et al.* [21] report on a computational method for DDIs prediction based on the functional similarity of drugs. Yan *et al.* [22] predict the DDI types by integrating the drug chemical, biological and phenotype data to calculate the embedding features with the cosine similarity method.

The feature-based methods [23–26] usually use drug features (such as molecular structure [27], targets [28], side effects [29], etc.) to represent drugs for DDI prediction. Feature selection, an important step, can filter out the most representative and relevant features. Based on DeepDDI [6], Deng *et al.* [30] develop a multimodal deep learning framework that integrates four drug features (substructures, targets, enzymes and pathways) as input to predict DDI events. Lin *et al.* [31] propose a supervised contrastive learning-based method for obtaining drug latent features that are more powerful for classification. However, feature-based methods rely heavily on expert experience and expertise.

Zhe Li is a graduate student at School of Computer, University of South China. He is interested in bioinformatics and graph neural networks.
Xinyi Tu is a graduate student at School of Computer, University of South China. Her research fields include drug discovery through deep learning methods.
Yuping Chen is a full professor at School of Pharmacy, University of South China. She won her Ph.D. in 2000 from Chinese Academy of Medical Sciences and Peking Union Medical College. Her research is to discover new molecular targets and drugs and to develop novel therapeutic delivery system.
Wenbin Lin is a full professor at School of Mathematics and Physics, University of South China. His research fields include deep learning and reinforcement learning.

Received: May 20, 2023. **Revised:** August 12, 2023. **Accepted:** September 13, 2023

The graph-based methods [32–35] usually organize DDI entries into a graph structure, where nodes are drugs and edges are interactions between drugs. Graph neural networks (GNNs) are the most adopted method for encoding drugs, e.g. graph convolutional network (GCN) [36, 37] and graph attention network (GAT) [38, 39]. Feng *et al.* [34] model the known DDIs as a signed and directed network and design a graph representation learning model to predict enhancive/depressive DDIs. He *et al.* [40] use 3D molecular graphs and position information to enhance the prediction ability for DDIs. In recent years, knowledge graphs (KGs) [41, 42] have attracted more attention due to their ability to provide comprehensive information. Lin *et al.* [8] learn drug representations with biomedical KG. Considering that the KG is large and noisy, Yu *et al.* [43] extract useful drug information from the local subgraph of KG. Su *et al.* [44] propose an attention-based KG representation learning framework, to fully utilize information of KGs for identifying potential DDIs effectively.

To improve the performance of DDI prediction, some recent works [45, 46] simultaneously take into account both feature-based and graph-based methods. Chen *et al.* [47] propose a two-layer intersect strategy that combines molecular graph and KG. Ren2022 [48] propose an inductive model to predict DDIs by aggregating local information and global information. Wang *et al.* [49] use a GCN with two DDI graph kernels to learn the feature representation of drugs. Pang *et al.* [50] develop an attention-based multi-dimensional feature encoder to process the Simplified Molecular-Input Line-Entry System (SMILES) string of drugs from 1D sequence and 2D graphical structure. Accordingly, It is beneficial for DDI prediction to fuse multi-scale features.

Although the above works have achieved encouraging results in DDI prediction, there are still some limitations: (i) Most previous works focus on if there exists an interaction between two drugs [21, 25, 51]. But in a clinic, predicting the specific interaction type is highly desirable. (ii) Most methods have performed well in identifying potential DDIs between known drugs [6, 43], but it is hard to predict interactions between new drugs. (iii) Generally, different entities have different types of connection [52], but homogeneous network-based methods treat all nodes equally.

In this paper, we propose a novel deep learning model based on a heterogeneous graph neural network (HGNN), named HetDDI, to predict unobserved DDIs. HetDDI takes into account the information not only from the molecular structure of drugs but also from the external biomedical KG. Firstly, we convert the drug SMILES into molecular graphs. The representation of molecular graphs is natural because graph-structured can reflect the properties of drugs. Secondly, we initialize the nodes embedding of graphs and related parameters by different pre-training methods, where the node-level method is used for the drug molecular graph and the link prediction method is used for the KG. Pre-training methods can enhance the generalization ability of the model. Then, we use two pre-trained HGNN blocks to learn the drug feature representation from molecular graphs and KG respectively. HGNN can effectively handle multiple nodes and types of hetero-graphs. Finally, we combine the above drug features and feed them into multi-layer perception (MLP) for binary-class, multi-class DDI and multi-label prediction tasks. In comparison, HetDDI has the following characteristics and advantages.

(i) **Improvement of the model's generalization ability.** Pre-training methods can provide a better initial representation, enabling the model with limited labeled data to perform well on DDI prediction between new drugs.

(ii) **Application of HGNN in extracting drugs' information.** HGNN-based approach can capture the structural information of drug molecular graphs and rich semantic information of biomedical KG, which significantly improves the performance of DDI prediction.

(iii) **Integration of drugs' intrinsic and extrinsic features.** HetDDI effectively integrates the features obtained from multi-source information, including drugs' molecular structure and biomedical KG, and confirms the importance of the synergistic effect of drugs' internal and external features for the downstream prediction tasks.

MATERIALS AND METHODS

Problem settings

A heterogeneous graph is defined as $G = (V, E)$. Here $V = \{v_1, v_2, \ldots, v_n\}$ is a set of n nodes, $E = \{e_1, e_2, \ldots, e_m\}$ is a set of m edges. A hetero-graph is represented by two mapping functions φ and ψ. $\varphi : V \rightarrow R_v$, which maps each node v to the corresponding type $\varphi(v)$, with R_v representing the set of node types. $\psi : E \rightarrow R_e$, which maps each edge e to the corresponding type $\psi(e)$, with R_e representing the set of edge types. The hetero-graph implies that $|R_v| > 1$ and / or $|R_e| > 1$.

There are N drugs in total. The drug set is defined as $D = \{d_1, d_2, \ldots, d_N\}$, and the drugs' 2D molecular structure graph set is defined as $G_{drug} = \{g_1, g_2, \ldots, g_N\}$. Here $g_i, i \in (1, N)$ is regarded as a hetero-graph, where atoms are nodes of the graph and chemical bonds are edges of the graph.

KG is generally expressed in the form of triplet, and we defined it as $G_{kg} = \{\mathfrak{h}, \mathfrak{r}, \mathfrak{t}\}$, where $\mathfrak{h}, \mathfrak{t} \in V$ that represents the entity set in KG. $\mathfrak{r} \in E$ which represents the set of relations. For ith triplet, $\mathfrak{r}_i$ represents the relation between entity $\mathfrak{h}_i$ and entity $\mathfrak{t}_i$.

In this paper, the DDI prediction task is to develop a computational model that takes drug pairs as input and achieve binary-class, multi-class and multi-label prediction task. For the binary-class and multi-label prediction task, we define an output matrix Y, where $Y_{ij} \in \{0, 1\}$, if $Y_{ij} = 1$ indicates that there is an interaction between drug pair (d_i, d_j), otherwise there is no interaction. For the multi-class prediction task, a total of 86 drug interaction relation types are defined [6]. Moreover, the drug data samples are divided into different types of training and test sets to evaluate the performance of HetDDI on three different scenarios, which are shown as follows (Figure 1):

(i) S1: to predict potential interactions between the known drugs.
(ii) S2: to predict potential interactions between known drugs and new drugs.
(iii) S3: to predict potential interactions between new drugs.

Overview of HetDDI

Figure 2 shows the framework of HetDDI. We first convert the sequence data into a graph structure and initialize the node embeddings by pre-training to enhance the model's generalization ability. Then HGNN is utilized to learn and iterate the vector representations of the drug chemical atoms and the KG nodes as well as their associated neighborhood entities. Finally, the classifier module concatenates the above-learned drug feature vectors and feeds them into three fully connected layers. The outputs are concatenated and sent to MLP with Dropout to predict the binary-class, multi-class and multi-label of DDIs on three scenarios.

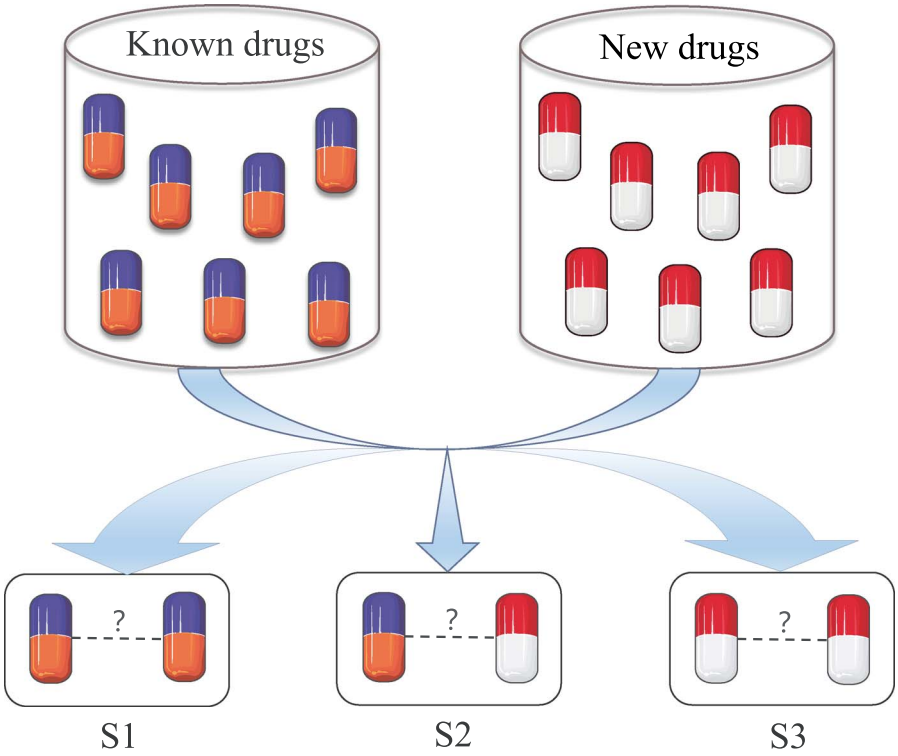

Figure 1. Problem definition. The known drugs and their known interactions attend the training. The new drugs refer to drugs that only appear in the test set, which are used to detect the performance of the trained model in predicting DDIs on S2 and S3. Parts of the figure were drawn by using pictures from Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

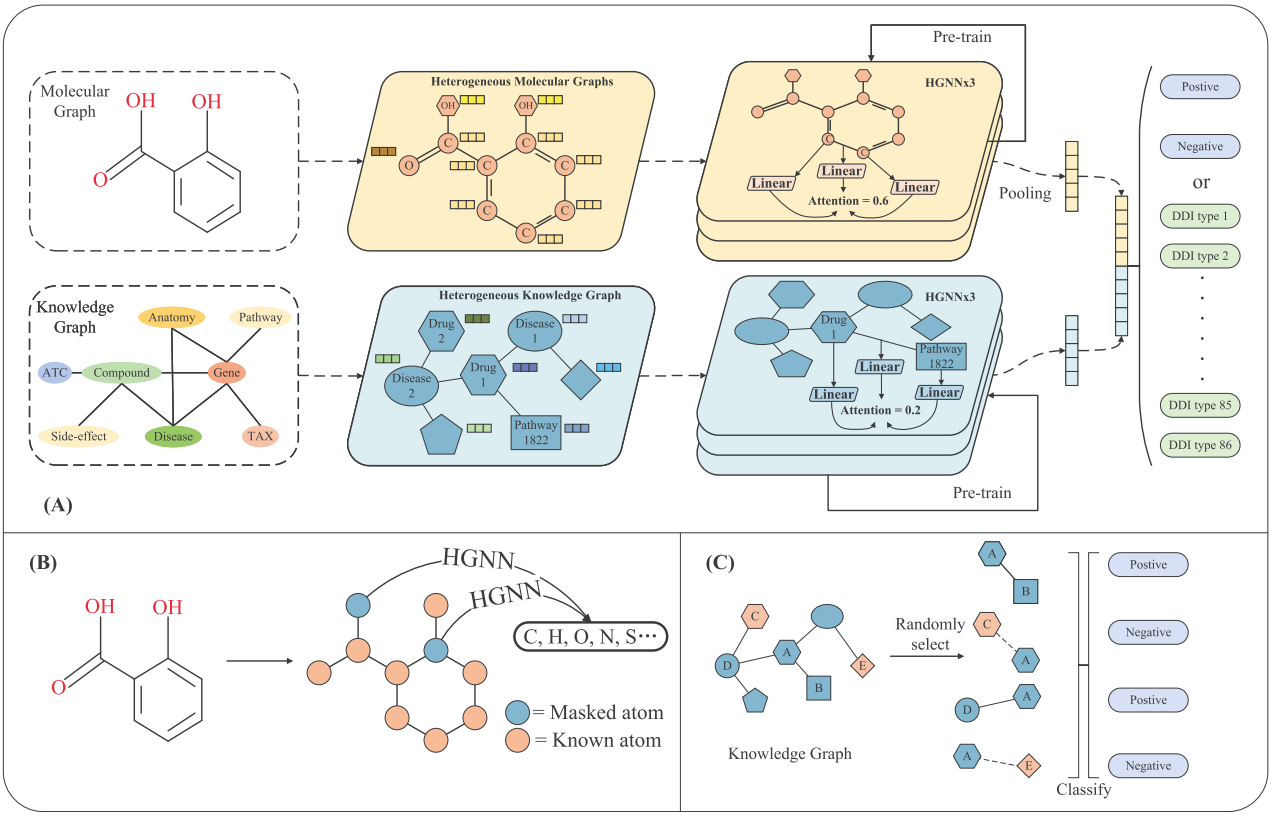

Figure 2. (**A**) The workflow of HetDDI: we transform the drug molecular and KG into hetero-graphs, and the initial HGNN parameters and drug feature embeddings are generated by pre-training. The pre-trained HGNN updates the drug feature vector representation by recursively aggregating heterogeneous information from adjacent nodes and edges. Finally, the two representations of drugs are concatenated and sent to the classifier for DDI prediction. (**B**) Pre-training on the drug molecule graph: we randomly mask the atomic types and make the network predict them. (**C**) Pre-training on the KG: we randomly select positive and negative samples in KG and perform link prediction to initialize network parameters.

Graph initialization and pre-training

To obtain richer drug structure information and semantic information, we regard both drug molecular graphs and KG as heterogeneous graphs:

(i) **Drug molecular graphs.** We treat each molecular graph as a hetero-graph during the initialization process, with atoms converted into nodes and chemical bonds corresponding to edges in the hetero-graph. Additionally, each type of atom (e.g. C, O, N, etc.) and chemical bond (e.g. single bond, double bond, etc.) has a unique representation, and it is through a randomly initialized embedding layer as a 1x300 vector. Therefore, a 38x300 atom embedding matrix and a 4 x 300 chemical bond embedding matrix are obtained.

(ii) **Knowledge graph.** Similarly, each entity in the KG (such as drug, target, disease or other biological entities) is considered as a node in the hetero-graph, and the relation between two entities is transformed into an edge in the hetero-graph. Unlike the molecular graph representation, each entity and relation has a unique 1 x 300 representation vector obtained through an embedding layer. Finally, a 97 243 x 300 entity embedding matrix and a 108 x 300 relation embedding matrix are formed.

We adopt two different pre-training methods to improve the generalization ability of the prediction model. For the drug molecular graph, we apply the node-level pre-training method [53] based on attribute masking that can pre-train the node embedding layer and the parameters of HGNN. As shown in (B) part of Figure 2, the type of 15% atoms in the drug molecule is randomly masked, which is used to predict the original types by HGNN and a simple linear model. The pre-training enables the HGNN to learn implicit chemical rules like the valence and electron of functional groups for further DDI prediction.

For the KG, follow the illustration of part (C) in Figure 2. We perform link prediction by sampling positive and negative samples, where the negative samples are from randomly replacing the tail of positive samples. In each training batch, we randomly select positive samples from the real dataset and an equal number of generated negative samples for training and remove the corresponding edges of positive samples in the KG. To pre-train parameters of the HGNN, we also use a linear layer to perform binary classification of the positive and negative samples, which promotes HGNN to achieve higher generalization ability in the subsequent tasks.

Heterogeneous graph neural network

A heterogeneous graph contains various types of nodes and edges, each with different attributes. In this paper, we apply HGNN to learn the drug's final representations by aggregating the drug molecular structure information and rich related information in the KG.

To address the impact of different node and edge features on the modeling results of a heterogeneous graph, we add an attention mechanism on HGNN, which considers different attribute information (e.g. molecular structure of drugs, the relations between drugs and diseases, and the chemical bonds between atoms, etc.) in the calculation of attention score. Thus, we assign different weights to neighbor edges during central node information aggregation. In each layer, we assign a d-dimension embedding for each neighbor edge type $\psi(e) \in R_e$ of the central node. In the lth layer, we learn the new embedding with a linear transformation matrix by connecting the node and edge feature vectors in a series and then apply a nonlinear activation function to obtain the attention score α_{ij}. Taking nodes i and j as an example, the detailed process is described as follows:

$$\alpha_{ij} = \sigma\left(A^T\left[W_{\varphi(i)}h_i^{(l-1)} \| W_{\varphi(j)}h_j^{(l-1)} \| W_{\psi(\langle i,j\rangle)}r_{\psi(\langle i,j\rangle)}\right]\right), \quad (1)$$

where $h_v^{(l-1)}$ is a feature representation vector for node v after $(l-1)$th layer. $\varphi(v)$ represents the type of node v, and $R_v = \{\varphi(v) : \forall v \in V\}$. $\psi(e)$ represents the type of edge e, and $R_e = \{\psi(e) : \forall e \in \varepsilon\}$, i.e., $\psi(\langle i,j\rangle)$ denotes the edge type between nodes i, j. $r_{\psi(\langle i,j\rangle)}$ is a feature representation vector of edge type $\psi(\langle i,j\rangle)$. A is a learnable parameter, $W_{\varphi(v)}$ and $W_{\psi(\langle i,j\rangle)}$ are learnable matrixs that maps features into embeddings. Superscript "T" means transposition. "$\|$" denotes a concatenation operation. σ is a nonlinear activation function (LeakyReLU is used here).

GNNs struggle to converge due to over-smoothing and gradient extinction problems [54, 55]. Recent studies have shown that the well-designed residual connection can partially solve the problem [56, 57]. Therefore, we add it that help the neural network design deeper to improve the fitting ability of the HGNN. For the lth layer, the aggregation process can be expressed as

$$h_i^{(l)} = ELU\left(h_i^{(l-1)} + \sum_{j \in N_i} \alpha_{ij} W_{\varphi(j)} h_j^{(l-1)}\right), \quad (2)$$

where ELU is an activation function.

Classifier module

Drug structure representation extraction. After k iterations of HGNN, the representation s_a of the atom a contains the structural information of its neighbor. Therefore, a graph-based single drug molecular representation s_i can be obtained by pooling all atom representations via the AVERAGE function, which can be described as:

$$s_i = AVERAGE\left(\left\{s_a^{(k)} | a \in g_i\right\}\right), \quad (3)$$

where AVERAGE represents average pooling, and $s_a^{(k)}$ denotes the representation of atom a at the k-th iteration.

Drug feature representation fusion and DDI prediction. After obtaining different drug representations, we concatenate the drug molecular structure-based representation s_i and the KG-based representation g_i as the drug d_i's final representation

$$f_i = [s_i \| g_i]. \quad (4)$$

For three different DDI prediction tasks, we concatenate two drugs' representations and send them into the MLP to predict the DDI possibility of drug pairs. In the binary-class and multi-label task, the process is described as follows:

$$\hat{y}_{ij} = sigmoid\left(MLP\left([f_i \| f_j]\right)\right), \quad (5)$$

where $\hat{y}_{ij}$ represents the probability score of interaction between the input drug pairs. In the multi-class prediction task, the process is described as follows:

$$\hat{y}_{ij} = softmax\left(MLP\left([f_i \| f_j]\right)\right), \quad (6)$$

where $\hat{y}_{ij}$ represents the probability score of each DDI type. Finally, in order to strengthen the generalization ability of the model, we

Table 1: The performance of HetDDI, its variants and six baselines on S1

Dataset	DrugBank									
tasks	Binary-class					Multi-class				
metrics	Accuracy	Precision	Recall	F1	AUC	Accuracy	Precision	Recall	F1	Kappa
DeepDDI	91.12	89.89	92.91	91.37	97.27	85.56	90.54	81.11	72.77	82.22
GAT	87.54	87.87	87.10	87.48	94.66	77.06	58.75	76.82	61.09	72.93
KGNN	92.75	92.99	92.98	92.97	97.31	92.58	79.94	73.77	75.92	91.17
MUFFIN	96.69	96.34	97.08	96.71	99.47	96.96	94.53	92.38	93.08	96.54
Molormer	97.05	96.32	97.91	97.11	99.67	96.77	94.87	92.45	93.91	96.17
MDF-SA-DDI	93.59	92.64	94.23	93.44	98.48	93.13	95.97	88.17	91.29	92.94
HetDDI-Mol	98.37	98.05	98.71	98.38	99.78	97.58	94.59	96.31	95.48	97.13
HetDDI-KG	98.72	98.42	99.04	98.73	99.85	97.96	94.74	96.54	95.86	97.57
HetDDI-UnPre	98.65	98.63	98.69	98.66	99.92	98.07	95.56	**96.72**	95.88	97.70
HetDDI-GIN	98.30	97.85	98.77	98.31	99.81	97.14	**96.81**	95.34	95.73	96.60
HetDDI (ours)	**98.82**	**98.52**	**99.12**	**98.82**	**99.87**	**98.13**	96.04	96.27	**96.17**	**97.78**

Note: Average values of 5-fold are shown in the table. For these metrics, higher values indicate better performance, and the best results are highlighted in bold.

Table 2: The performance of HetDDI, its variants and four baselines on S1

Dataset	Twosides				
tasks	multi-label				
metrics	Accuracy	Precision	Recall	F1	AUC
DeepDDI	87.78	86.63	89.30	87.94	94.61
KGNN	92.09	93.30	90.71	91.99	97.55
MUFFIN	95.18	93.42	97.20	95.28	98.88
Molormer	94.81	92.40	97.60	94.93	98.74
HetDDI-Mol	96.05	94.39	97.11	96.12	99.15
HetDDI-KG	96.18	95.12	97.26	96.23	99.18
HetDDI-UnPre	96.32	95.38	**97.36**	96.36	99.23
HetDDI-GIN	96.20	95.40	97.09	96.24	99.19
HetDDI (ours)	**96.66**	**96.08**	97.29	**96.68**	**99.34**

Note: Average values of five-fold are shown in the table. For these metrics, higher values indicate better performance, and the best results are highlighted in bold.

add the Dropout function to the full connection layer of MLP and present the experimental results in Table 1 and Table 2.

Training

During training, we optimize the model parameters by minimizing cross-entropy loss to improve multi-task DDI prediction results. In the binary-class and multi-label prediction task, for the DDI triplet (d_i, r_{ij}, d_j), the loss function is defined as:

$$LOSS_b = \sum_{(d_i, r_{ij}, d_j) \in N_d} -y_{ij} \log \hat{y}_{ij} - (1 - y_{ij}) \log (1 - \hat{y}_{ij}), \quad (7)$$

where N_d is the set of all DDI triplets. r_{ij} represents that there is an interaction in the binary-class task or one of the 200 label types in the multi-label. $y_{ij} \in \{0, 1\}$ indicates whether the (d_i, r_{ij}, d_j) is positive or negative sample. If $y_{ij} = 1$, it indicates that the triplet is positive. $\hat{y}_{ij}$ is the predicted probability of triplet (d_i, r_{ij}, d_j) is positive.

In the multi-class prediction task, the loss function is defined as:

$$LOSS_m = -\sum_{(d_i, d_j) \in N_d} \sum_{t=1}^{N_t} y_{ij}^{(t)} \log \hat{y}_{ij}^{(t)}, \quad (8)$$

where $N_t = 86$ is the total number of drug interaction types. $y_{ij}^{(t)} \in \{0, 1\}$ denotes whether the actual condition of the current drug pair (d_i, d_j) belongs to type t, and $\hat{y}_{ij}^{(t)}$ is the predicted interaction probability belonging type t for the drug pair (d_i, d_j).

EXPERIMENTS

We demonstrate the performance of HetDDI in predicting DDIs via several comparisons and ablation experiments.

Datasets and settings

Datasets. (i) **DrugBank** [58] is a large-scale drug knowledge database, which provides chemical structure, pharmacology, drug action and other comprehensive data of over 50 000 drugs and their derivatives. The dataset contains 1710 drugs and 192 284 known DDIs with a total of 86 relation types [6], and the complete list of types is shown in Supplementary Table S3. We get SMILES of all drugs from DrugBank, and we discard SMILES that cannot be expressed as a molecular graph and remove their associated relations. Finally, there are 1706 drugs and 191 427 DDIs for our experiments. We randomly select drug pairs without interactions in the dataset as negative samples of the binary-class task, and we balance the ratio of positive and negative samples to 1:1.

(ii) **TWOSIDES** [59] collects a large amount of drug-related side effect information, including 3300 drugs and 42 920 392 DDIs with 12 710 interaction types. It allows for multiple interactions

between a given drug pair, thus the dataset is used for multi-label DDI prediction. To establish a map between drugs and KG, we remain 1 979 575 DDIs with 200 interaction types, ensuring a focus on the interaction types that occur around 10 000 times, and the specific types are shown in Supplementary Table S4. The negative samples are generated by replacing d_i or d_j in the known DDI triplet (d_i, r_{ij}, d_j) and the number of negative samples is the same as positive samples.

(iii) **KG Dataset.** We use DRKG [60] as the external biomedical KG, which consists of 97 238 entities belonging to 13 entity types and 5 874 261 triplets belonging to 107 edge types. Furthermore, we delete the same DDIs in the KG as in the DrugBank or TWOSIDES to prevent data leakage issues.

Baselines. We compare HetDDI with the following state-of-the-art models:

- DeepDDI [6] is a deep learning model. It takes the drug pair and their structural information as input to predict multi-class DDIs and the food-drug interaction effects.
- GAT [38] introduces the attention mechanism to assign different weights for each node according to their importance, which helps the model learn structural information in the DDI network.
- KGNN [8] proposes a KG-based GNN model that explores structural and semantic information in the KG for potential DDI prediction.
- MUFFIN [47] is a multi-scale feature fusion deep learning framework, which designes a bi-level cross strategy to fuse multi-modal features well.
- Molormer [61] applies attention-based methods to process the molecular graph encoded by spatial information.
- MDF-SA-DDI [62] predicts DDI events based on different multi-source drug fusion methods and different multi-source feature fusion methods.

To verify the necessity of each module and method in our model, we design several variants of HetDDI:

- HetDDI-Mol only takes into account of drug's molecular graph structure information to make DDI prediction.
- HetDDI-KG only considers external medical KG information to make DDI prediction.
- HetDDI-UnPre makes DDI prediction without using pre-training methods.
- HetDDI-GIN replaces the HGNN module with the ordinary graph isomorphic network (GIN), and keeps other parameters unchanged.

Metrics. The prediction results can be divided into True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) in the confusion matrix. We evaluate the performance of HetDDI via the following metrics that are widely used in the classification task.

- Accuracy: the proportion of predicting numbers correctly, $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$.
- Precision: the proportion of true positive in the predicted positive samples, $Precision = \frac{TP}{TP+FP}$.
- Recall: the proportion of all positive samples that are predicted correctly, $Recall = \frac{TP}{TP+FN}$.
- F1-score: a weighted average of precision and recall, $F1 = \frac{2\times Precision\times Recall}{Precision+Recall}$
- ROC–AUC: the area under the ROC curve. The vertical axis is $\frac{TP}{TP+FN}$, and the horizontal axis is $\frac{FP}{TN+FP}$. Assuming that the ROC curve is a sequence of points with $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, then $ROC\text{-}AUC = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i+1} - x_i)\cdot(y_{i+1}+y_i)$.
- Kappa: a consistency test indicator, defined as $kappa = \frac{p_o - p_e}{1-p_e}$, where p_o is the classification accuracy of each DDI type, and p_e is the ratio of the sum for the sample number multiplied by the predicted one in each class to the square of the total sample number.

Experimental settings. The learning rate is set as 0.0001. The drug molecular structure and KG information are represented by 300-dimensional vectors through HGNN. In order to enhance the generalization ability of HetDDI, we set the weight decay coefficient as 0.0002, the dropout rate of the classifier as 0.5, and add batch regularization. According to different task requirements, the number of neurons set in the output layer is also different: one for the binary-class task, 86 for the multi-class one and 200 for multi-label.

In addition, we set different experiments for DDI prediction on three scenarios. For S1, we randomly divide all DDIs into five parts, where four of them as a training set and the remaining one as a test set. It is a transductive learning process in that we train the model based on the training set and predict DDIs in the test set. For S2 and S3, we randomly split drugs into two sets in the same ratio of 4:1, with one set being a training set including known drugs and the other set being a test set containing new drugs. We use an inductive learning approach to train the model on DDIs between two known drugs. S2 predicts the DDIs between one known drug and one new drug, while S3 predicts the DDIs between two new drugs. Meanwhile, all experiments are based on five-fold cross-validation. To address the class imbalance, we employ an up-sampling technique in the training set to expand the number of DDI types with a smaller sample size.

Experimental results and analysis

Results and analysis on the transductive scenario

The upper part of Table 1 and Table 2 report the predictive performance of HetDDI and baseline methods for binary-class, multi-class, and multi-label DDI prediction tasks on S1. On the DrugBank dataset, HetDDI exhibits the best performance, where the accuracy achieves 98.82% in the binary-class task and 98.13% in the multi-class one. Molormer has a good performance because it fully considers the spatial structure information of drugs. Our model considers not only drug molecular structure but also the rich information in the KG, which outperforms Molormer by 1.7% and 2.3% in terms of F1-score on the two tasks respectively. These demonstrate the advantage of HetDDI in predicting DDIs between two known drugs.

On the Twosides dataset, the accuracy, precision, F1-score and AUC have improved at least 1.5%, 2.66%, 1.4% and 0.5% for the multi-label task. Additionally, MUFFIN performs best among all baselines because it fuses multi-scale features of drugs well. The reason for the super performance of HetDDI is that it is based on the success of MUFFIN, employing HGNN to further learn the multi-source heterogeneous information to synergize the DDI prediction.

Results and analysis on the inductive scenario

Compared Table 1 with Table 3, the scores of all evaluation metrics on S2 and S3 are significantly lower than those on S1, which indicates that the inductive task is more challenging than the transductive task. The prediction difficulty gradually increases across these three scenarios due to new drugs only existing in the test set. We apply HGNN to pre-train the two data sources for improving the generalization ability of the model. As shown in the

Table 3: The performance of multi-class prediction task on S2 and S3 among HetDDI, its variants and four baselines

Task	Multi-class									
scenarios	S2 (known and new drugs)					S3 (new drugs)				
metrics	Accuracy	Precision	Recall	F1	Kappa	Accuracy	Precision	Recall	F1	Kappa
DeepDDI	61.97	54.61	46.63	53.63	58.86	40.89	30.60	25.82	22.92	37.61
MUFFIN	74.04	69.53	62.45	63.19	72.96	51.96	34.99	23.57	25.77	44.73
Molormer	62.34	47.58	37.56	40.15	57.17	40.21	24.00	16.89	17.47	36.78
MDF-SA-DDI	66.11	57.77	53.13	52.71	60.36	46.54	34.35	27.79	22.89	40.31
HetDDI-Mol	76.91	60.79	61.11	58.49	72.29	61.87	40.13	33.22	34.49	53.95
HetDDI-KG	78.79	63.44	62.42	58.67	75.88	64.23	40.86	36.45	35.42	58.34
HetDDI-UnPre	80.64	68.98	68.55	63.88	75.86	65.51	42.26	38.97	37.85	58.84
HetDDI-GIN	79.08	65.64	63.09	60.64	77.52	64.58	42.82	36.91	36.06	58.75
HetDDI (ours)	**81.93**	**71.08**	**70.51**	**68.20**	**78.16**	**66.56**	**44.15**	**40.13**	**40.01**	**60.74**

Note: Average values of 5-fold are shown in the table. The best results are highlighted in bold.

Table 4: The performance of different type combinations of nodes in the KG on S3

Node type	Accuracy	Precision	Recall	F1	Kappa
C	60.15	32.17	27.87	27.39	52.10
C+G	63.51	39.96	31.70	32.95	54.85
C+D	61.83	35.42	32.47	32.56	54.57
C+S	61.59	32.01	29.93	29.11	54.39
C+A	61.16	31.59	29.52	28.79	53.92
C+P	60.59	28.62	26.50	24.91	50.27
C+G+D	63.65	40.35	36.16	35.01	57.88
C+G+D+S+A+P	63.91	36.38	**36.58**	34.76	54.46
HetDDI-KG (all types)	**64.23**	**40.86**	36.45	**35.42**	**58.34**

Note: Average values of 5-fold are shown in the table. For these metrics, higher values indicate better performance, and the best results are highlighted in bold.

upper part of Table 3, HetDDI acquires exciting results compared with other state-of-the-art methods.

Specifically, HetDDI achieves an accuracy of 81.93% and an F1-score of 68.20% on S2, outperforming the best baseline by up to 7.89% on accuracy and 5.01% on F1-score. For S3, the accuracy and F1-score have been improved by at least 14.60% and 14.24%, respectively, which particularly demonstrates the superiority of HetDDI. Because we make up for the lack of internal knowledge of new drugs by learning knowledge from two data sources.

Ablation study

To explore the effect of every module of our model, we further contrast HetDDI with its four variants on two datasets, and the comparison results are shown in the lower part of Table 1, Table 2 and Table 3. We can observe that our model performance improves by an average of 2% on the transductive scenario and 5% on the inductive scenario. In particular, the application of HGNN offers great help to our model.

On the transductive scenario, Table 1 shows that multi-source information fusion has an inapparent improvement on the final results due to the single approach (HetDDI-Mol or HetDDI-KG) is already beneficial for learning useful drug representations. On the Twosides dataset, HetDDI has a slight improvement in the multi-label task compared with other variants. Additionally, HetDDI outperforms HetDDI-Mol and HetDDI-KG in terms of all metrics on three different tasks, thus further supporting the effectiveness of aggregating multi-source information.

On the inductive scenario, HetDDI has significant improvement in results on the multi-class prediction task. Taking the F1-score as an example, HetDDI-Mol utilizes the molecular graphs to capture drug features, while ignoring the KG information, resulting in a significant reduction of 9.71% and 5.52% for S2 and S3 respectively. HetDDI-KG considers topological structures and semantic information in the KG to learn the feature representation of entities but does not incorporate drug molecular structure information, resulting in a 9.53% and 4.59% decrease for S2 and S3 respectively. HetDDI-UnPre directly predicts interactions between new drugs without relying on pre-training methods, and the F1 score achieves a reduction of 4.32% on S2 and 2.16% on S3. Because pre-training can help the model to learn more general and meaningful representations of drug features from a large amount of unlabeled data, which improves the generalization ability of the model for unknown drugs. Additionally, HetDDI-GIN uses a drug graph isomorphic network without node and edge type information, and its performance is 7.56% and 3.95% lower than HetDDI for S2 and S3, respectively. In summary, the aggregation of muti-source heterogeneous information can provide rich drug information from different views to predict DDI events.

Effects of different types of nodes

Graph learning on heterogenous graphs is very challenging, so it is crucial to investigate how nodes of different types contribute to the improved performance. The KG contains 13 types of nodes, but only five types of Gene (G), Disease (D), Side-effect (S), ATC (A) and Pharmacologic class (P) are directly connected to Compound (C) and have abundant information on their interrelation. The different types of biological nodes mainly provide potential feature information when predicting DDIs between new drugs. We systematically conduct experiments on S3 to verify the effects of each type on our model's performance.

As shown in Table 4, we observe that some types of nodes play pivotal roles in specific aspects of DDI prediction. For example,

Diseases and Side-effects are crucial indicators of drug interactions and safety. The accuracy of C+G reaches 63.51%, a remarkable improvement over those without Gene. This can be attributed to the fact that Genes can enhance our model's ability to capture target information associated with drugs. The absence of Gene nodes will make the model lose important information about drug interactions at the molecular level. It is also worth noting that the performance of the model gradually improves as more types are taken into account. This trend suggests that nodes of types in the KG can provide auxiliary and complementary information for predicting DDIs on S3.

Overall, the nodes of different types in the KG have varying degrees of impact on our model's performance. The diversity of types enables HetDDI to take full advantage of rich biological information, leading to the accurate DDI predictions. This further demonstrates the importance of considering both molecular structure of drugs and biological KG in our model.

Parameter analysis

We study the effect of some important parameters on model performance. When evaluating one parameter, we fix all other parameters. Figure 3 shows the effect of different GNN layers and embedding dimensions on the experimental results.

Effect of GNN layers. Parts (1) and (2) of Figure 3 show the effect of different GNN layer number l (from 1 to 5) on the model performance. It is found that the performance of HetDDI starts to decline from layer=3, because a larger layer number will make the model overfit, and a smaller layer number will cause underfitting. So we adopt a three-layer GNN in HetDDI to achieve the best results.

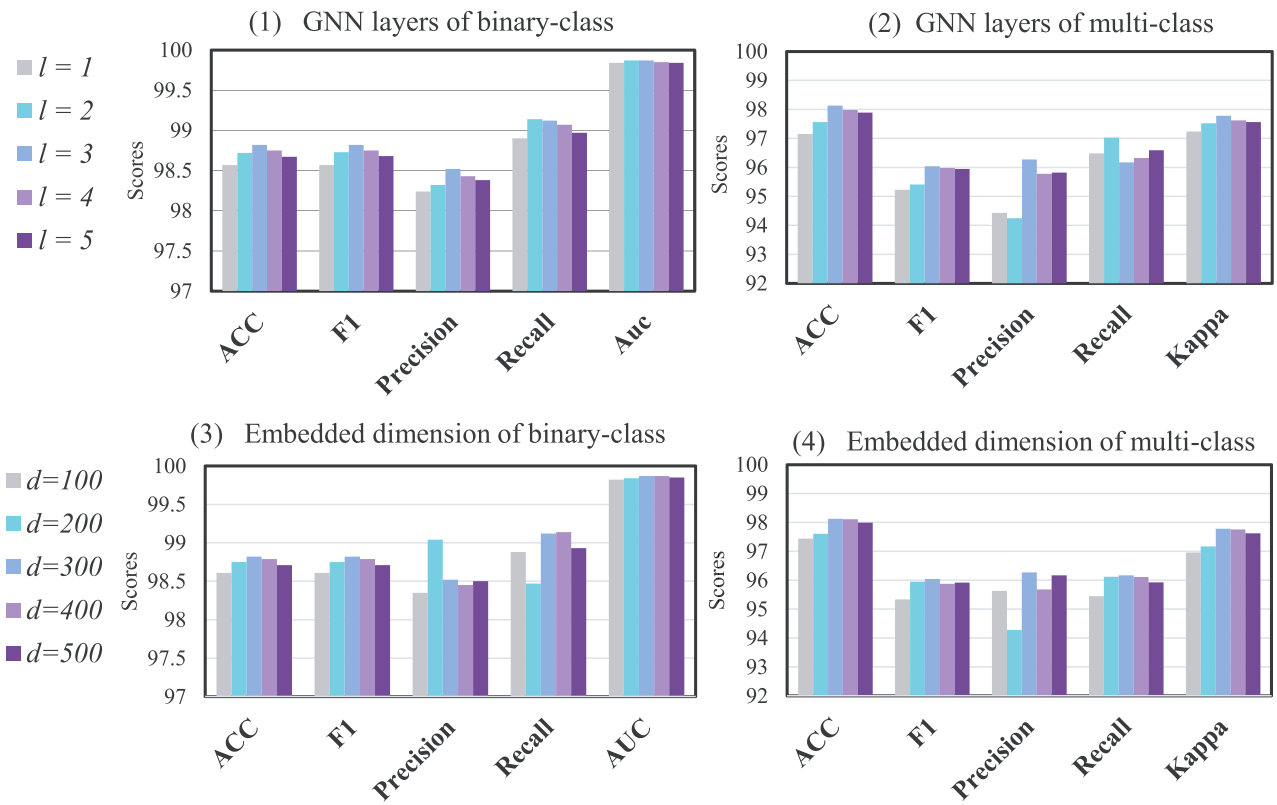

Figure 3. The performance of HetDDI under different GNN layers and embedding dimension settings.

Table 5: The predicted top 10 DDIs in DrugBank and the corresponding proofs

Drug1 ID	Drug2 ID	Drug–drug interaction	Proof
DB00564	DB00705	The metabolism of Drug2 can be increased when combined with Drug1	(1)
DB00333	DB00834	The risk or severity of QTc prolongation can be increased when Drug1 is combined with Drug2	(2)
DB09118	DB09065	The metabolism of Drug2 can be decreased when combined with Drug1	(3)
DB00252	DB00243	The metabolism of Drug2 can be increased when combined with Drug1	(4)
DB01065	DB08820	The serum concentration of Drug2 can be increased when combined with Drug1	No proof
DB00312	DB09065	The metabolism of Drug2 can be increased when combined with Drug1	(5)
DB01320	DB00243	The metabolism of Drug2 can be increased when combined with Drug1	(6)
DB01174	DB00289	The metabolism of Drug2 can be increased when combined with Drug1	No proof
DB01211	DB00834	The metabolism of Drug2 can be decreased when combined with Drug1	(7)
DB08820	DB09034	The serum concentration of Drug2 can be increased when it is combined with Drug1	(8)

Note: The proofs (1)–(8) can be found in the following websites. (1) https://www.drugs.com/interactions-check.php?drug_list=497-0,794-0 (2) https://go.drugbank.com/drug-interaction-checker (3) https://go.drugbank.com/drug-interaction-checker (4) https://go.drugbank.com/drug-interaction-checker (5) https://go.drugbank.com/drug-interaction-checker (6) https://go.drugbank.com/drug-interaction-checker (7) https://go.drugbank.com/drug-interaction-checker (8) https://go.drugbank.com/drug-interaction-checker.

Effect of embedding dimensions. Parts (3) and (4) of Figure 3 show the effect of different embedding dimensions d (from 100 to 500) on the model performance. It is found that information can be well represented by appropriately increasing the embedding dimensions. However, larger embedding dimensions will increase the complexity of the model and cause overfitting. So we adopt 300 as the embedding dimension to achieve optimal performance.

Case study

We demonstrate the predictive ability of HetDDI through case study. We train the model using all known DDI events from the DrugBank dataset and then make potential predictions for other unobserved DDIs. We sequentially converted the prediction scores into a list of recommendations for unknown DDIs, where a higher score for a drug pair indicates a greater likelihood of interactions. Table 5 lists the top 10 drug pairs and their predicted DDI types. Finally, we use drugs.com and the Drug Interaction Checker tool provided by DrugBank to find evidence to verify whether these predictions are accurate.

We can observe that 8 of the top 10 drug pairs are supported by evidence. For instance, Ranolazine (DB00243) is mainly metabolized by the CYP3A4 enzyme [63]. However, when administered concurrently with Phenytoin (DB00252), a CYP3A4 inducer, it can lead to increased metabolism and decreased serum concentrations of ranolazine, reducing anti-anginal efficacy [64]. All the case studies further demonstrate the strong predictive performance of HetDDI.

CONCLUSION

In this work, we propose a novel HGNN-based DDI prediction model, which provides an effective method for learning multi-source drug features from drug molecular graphs and KG by exploiting the multi-relation information of hetero-graphs. Experimental results demonstrate that HetDDI has better prediction performance on both three tasks and scenarios than the state-of-the-art baseline models. Therefore, HetDDI is useful in preventing the occurrence of adverse events in clinical drug treatment.

Currently, we focus on applying HGNN to DDI prediction tasks, we will try to provide an interpretable prediction for DDIs by improving the attention mechanism in the future. Furthermore, the model can also be generalized to other tasks, such as the prediction of drug-target interactions.

Key Points

- We propose an HGNN-based model (HetDDI), which considers the heterogeneous information not only from drug molecular graphs but also from biomedical KG.
- The heterogeneous graph neural network (HGNN) initialize parameters via pre-training to enhance the generalization ability of the model.
- Esxperiments confirm the effectiveness of our model on three different tasks.
- The model has achieved promising performance compared with the state-of-the-art methods on the transductive and inductive scenarios.
- Case study further indicates our model's ability to predict unobserved DDIs.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bib.oxfordjournals.org/.

FUNDING

This work was supported in part by Top Foreign Experts of the Ministry of Science and Technology of China (G2021029011L).

DATA AVAILABILITY

All data used in the study are from public resources. DrugBank is available at https://bitbucket.org/kaistsystemsbiology/deepddi/src/master/data/. TWOSIDES is available at https://tatonettilab.org/offsides/. DRKG is available at https://github.com/gnn4dr/DRKG/.

REFERENCES

1. Giacomini KM, Krauss RM, Roden DM, *et al*. When good drugs go bad. *Nature* 2007;**446**(7139):975–7.
2. Greger J, Bates V, Mechtler L, Gengo F. A review of cannabis and interactions with anticoagulant and antiplatelet agents. *J Clin Pharmacol* 2020;**60**(4):432–8.
3. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today* 2005;**10**(21):1421–33.
4. Gao H, Korn JM, Ferretti S, *et al*. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* 2015;**21**(11):1318–25.
5. Lin X, Quan Z, Wang Z-J, *et al*. A novel molecular representation with bigru neural networks for learning atom. *Brief Bioinform* 2020;**21**(6):2099–111.
6. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc Natl Acad Sci* 2018;**115**(18):E4304–11.
7. Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug–drug interaction effects. *BMC Bioinformatics* 2019;**20**(1):1–8.
8. Lin X, Quan Z, Wang Z-J, *et al*. Kgnn: knowledge graph neural network for drug–drug interaction prediction. *IJCAI* 2020;**380**:2739–45.
9. Huang L, Lin J, Li X, *et al*. Egfi: drug–drug interaction extraction and generation with fusion of enriched entity and sentence information. *Brief Bioinform* 2022;**23**(1):bbab451.
10. Hui Y, Zhao SY, Shi JY. Stnn-ddi: a substructure-aware tensor neural network to predict drug–drug interactions. *Brief Bioinform* 2022;**23**(4):bbac209.
11. Zhao Z, Yang Z, Luo L, *et al*. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics* 2016;**32**(22):3444–53.
12. Vilar S, Friedman C, Hripcsak G. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform* 2018;**19**(5):863–77.
13. Lim S, Lee K, Kang J. Drug drug interaction extraction from the literature using a recursive neural network. *PloS One* 2018;**13**(1):e0190926.
14. Asada M, Miwa M, Sasaki Y. Using drug descriptions and molecular structures for drug–drug interaction extraction from literature. *Bioinformatics* 2021;**37**(12):1739–46.

15. Hong L, Lin J, Li S, *et al*. A novel machine learning framework for automated biomedical relation extraction from large-scale literature repositories. *Nat Mach Intell* 2020;**2**(6):347–55.
16. Han K, Peigang Cao Y, Wang FX, *et al*. A review of approaches for predicting drug–drug interactions based on machine learning. *Front Pharmacol* 2022;**12**:3966.
17. Vilar S, Harpaz R, Uriarte E, *et al*. Drug—drug interaction through molecular structure similarity analysis. *J Am Med Inform Assoc* 2012;**19**(6):1066–74.
18. Li P, Huang C, Yingxue F, *et al*. Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015;**31**(12): 2007–16.
19. Yan C, Duan G, Pan Y, *et al*. Ddigip: predicting drug–drug interactions based on gaussian interaction profile kernels. *BMC Bioinformatics* 2019;**20**(15):1–10.
20. Rohani N, Eslahchi C. Drug–drug interaction predicting by neural network using integrated similarity. *Sci Rep* 2019;**9**(1): 13645.
21. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform* 2017;**70**:54–64.
22. Yan C, Duan G, Zhang Y, *et al*. Predicting drug-drug interactions based on integrated similarity and semi-supervised learning. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**19**(1):168–79.
23. Gottlieb A, Stein GY, Oron Y, *et al*. Indi: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 2012;**8**(1):592.
24. 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):e278–86.
25. Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PloS One* 2018;**13**(5):e0196865.
26. Kexin Huang, Cao Xiao, Trong Hoang, Lucas Glass, and Jimeng Sun. Caster: predicting drug interactions with chemical substructure representation. *In:Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. **34**, p. 702–9, 2020.
27. Zhang W, Chen Y, Liu F, *et al*. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics* 2017;**18**:1–12.
28. Takeda T, Hao M, Cheng T, *et al*. Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. *J Chem* 2017;**9**(1):1–9.
29. Zhang P, Wang F, Jianying H, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**(1):12339.
30. Deng Y, Xinran X, Qiu Y, *et al*. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 2020;**36**(15):4316–22.
31. Lin S, Chen W, Chen G, *et al*. Mddi-scl: predicting multi-type drug-drug interactions via supervised contrastive learning. *J Chem* 2022;**14**(1):1–12.
32. Huang K, Xiao C, Glass LM, *et al*. Skipgnn: predicting molecular interactions with skip-graph networks. *Sci Rep* 2020;**10**(1): 1–16.
33. Yao J, Sun W, Jian Z, *et al*. Effective knowledge graph embeddings based on multidirectional semantics relations for polypharmacy side effects prediction. *Bioinformatics* 2022;**38**(8):2315–22.
34. Feng Y-H, Zhang S-W, Feng Y-Y, *et al*. A social theory-enhanced graph representation learning framework for multitask prediction of drug–drug interactions. *Brief Bioinform* 2023;**24**(1):bbac602.
35. Li Z, Zhu S, Shao B, *et al*. Dsn-ddi: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Brief Bioinform* 2023;**24**(1):bbac597.
36. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. 2016.
37. Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. *In: Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. , p. 7370–7, 2019.
38. Velickovic P, Cucurull G, Casanova A, *et al*. Graph attention networks. *Stat* 2017;**1050**(20):10–48550.
39. Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: knowledge graph attention network for recommendation. In:*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, p. 950–958, 2019.
40. He H, Chen G, Chen CY-C. 3dgt-ddi: 3D graph and text based neural network for drug–drug interaction prediction. *Brief Bioinform* 2022;**23**(3):bbac134.
41. Fensel D, Şimşek U, Angele K, *et al*. Introduction: what is a knowledge graph? *Knowl Graphs: Methodol Tools Select Cases* 2020;**02**: 1–10.
42. Zhang J, Chen M, Liu J, *et al*. A knowledge-graph-based multimodal deep learning framework for identifying drug–drug interactions. *Molecules* 2023;**28**(3):1490.
43. Yue Y, Huang K, Zhang C, *et al*. Sumgnn: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* 2021;**37**(18):2988–95.
44. Xiaorui S, Lun H, You Z, *et al*. Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Brief Bioinform* 2022;**23**(3):bbac140.
45. Md Rezaul Karim, Michael Cochez, Joao Bosco Jares, Mamtaz Uddin, Oya Beyan, and Stefan Decker. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network. In:*Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, p. 113–123, 2019.
46. Han X, Xie R, Li X, Li J. Smilegnn: drug–drug interaction prediction based on the smiles and graph neural network. *Life* 2022;**12**(2):319.
47. Chen Y, Ma T, Yang X, *et al*. Muffin: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* 2021;**37**(17): 2651–8.
48. Ren Z-H, You Z-H, Chang-Qing Y, *et al*. A biomedical knowledge graph-based method for drug–drug interactions prediction through combining local and global features with deep neural networks. *Brief Bioinform* 2022;**23**(5):bbac363.
49. Wang F, Lei X, Liao B, Fang-Xiang W. Predicting drug–drug interactions by graph convolutional network with multi-kernel. *Brief Bioinform* 2022;**23**(1):bbab511.
50. Pang S, Zhang Y, Song T, *et al*. Amde: a novel attention-mechanism-based multidimensional feature encoder for drug–drug interaction prediction. *Brief Bioinform* 2022;**23**(1):bbab545.
51. Feng Y-H, Zhang S-W, Shi J-Y. Dpddi: a deep predictor for drug-drug interactions. *BMC Bioinformatics* 2020;**21**(1):1–15.
52. Liu S, Zhang Y, Cui Y, *et al*. Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;**23**(2):976–85.
53. Hu W, Liu B, Gomes J, *et al*. Strategies for pre-training graph neural networks. arXiv preprint arXiv:1905.12265. 2019.
54. Qimai Li, Zhichao Han, and Xiao-Ming Wu. *Deeper insights into graph convolutional networks for semi-supervised learning*. In:*Thirty-Second AAAI Conference on Artificial Intelligence*, 2018, **32**.
55. Keyulu X, Li C, Tian Y, *et al*. Representation learning on graphs with jumping knowledge networks. In:*International Conference on Machine Learning*. PMLR, 2018, 5453–62.

56. Li G, Xiong C, Thabet A, Ghanem B. Deepergcn: all you need to train deeper gcns. arXiv preprint arXiv:2006.07739. 2020.
57. He R, Ravula A, Kanagal B, Ainslie J. Realformer: transformer likes residual attention. arXiv preprint arXiv:2012.11747. 2020.
58. Wishart DS, Feunang YD, Guo AC, *et al*. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res* 2018;**46**(D1):D1074–82.
59. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**(125):125ra31.
60. Ioannidis VN, Song X, Manchanda S, *et al*. Drkg-drug repurposing knowledge graph for Covid-19. arXiv preprint arXiv: 2010.09600. 2020.
61. Zhang X, Wang G, Meng X, *et al*. Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug–drug interactions prediction. *Brief Bioinform* 2022;**23**(5):bbac296.
62. Lin S, Wang Y, Zhang L, *et al*. Mdf-sa-ddi: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief Bioinform* 2022;**23**(1): bbab421.
63. Jerling M. Clinical pharmacokinetics of ranolazine. *Clin Pharmacokinet* 2006;**45**:469–91.
64. Trujillo TC. Advances in the management of stable angina. *J Manag Care Pharm* 2006;**12**:S10–6.