Enhancing Drug-Drug Interaction Event Prediction from Knowledge Graphs by Multimodal Deep Neural Networks

Tengfei Lyu, Jianliang Gao, Ling Tian, Zhao Li, Peng Zhang, and Ji Zhang

Abstract—Drug-Drug Interaction (DDI) events prediction is a meaningful and challenging task in pharmacology and clinical application, and effective prediction of the DDI events could help clinicians make effective decisions and establish appropriate therapy programs. Recently, many deep learning techniques have been proposed for DDI events prediction. However, most existing methods pay less attention to the potential correlations between DDI events and other multimodal data. In this study, we propose a novel framework for DDI events prediction named **M**ultimodal **D**eep **N**eural **N**etwork (MDNN), which explores the features of drug multimodal data such as drug knowledge graph and heterogeneous features in a unified framework. In MDNN, we design a two-pathway framework including drug knowledge graph (DKG) based pathway and heterogeneous feature (HF) based pathway to obtain drug multimodal representations. Our framework can effectively capture drugs and their potential neighborhoods by mining their associated relations in DKG. And it also applied to effectively assist the joint representation learning of both structure information and the multimodel data. We have implemented our method and conducted experiments on a real-world dataset. Empirical results show that MDNN outperforms the classic and state-of-the-art models.

Index Terms—drug-drug interaction, DDI events prediction, graph neural network, multimodal

1 Introduction

With the rapid growth of the number of drug types, it is essential to manage drug safety when multiple drugs are adopted in the treatment of a disease [1], [2]. Drug-Drug Interactions (DDI) often occur in cases of simultaneous administration of multiple drugs, which may result in adverse drug reactions that cause injuries and huge medical costs [3], [4]. However, DDI can lead to different biological consequences and events. Different drug combinations produce different drug interaction events, and the body will exhibit different side effects [5]. For example, as shown in Figure 1, drug *Itraconazole* and drug *Abemaciclib* interaction together cause an event that *the risk to increase due to the severity of the adverse effects*. However, it will cause *the decrease of body's serum concentration* when mixing drug *Abemaciclib* and drug *Dabrafenib*. Therefore, accurate prediction of DDI events becomes a clinically important task that could help clinicians make effective decisions and establish appropriate therapy programs. The correct use of multiple drugs can minimize the medical risks while maximizing the synergy benefits of drugs [6].

Specifically, DDI events prediction can be seen the multi-class classification task which is a meaningful and challenging task and has received some attention. Predicting potential DDI events reduces unexpected drug-drug interactions, lowers drug development costs, and can be used to optimize the drug design process [7], [8]. To mitigate these risks and costs, accurate prediction of DDIs becomes a clinically important task [6]. Therefore, to alleviate the impact of unexpected pharmacological effects, it is an immediate need for finding an effective prediction model to solve the problem, which can minimize unexpected adverse drug reactions and maximize synergistic benefits when treating using multi-drugs [9].

There have been a number of deep learning models proposed for DDI events prediction, including analyzing chemical structure similarity using graph neural network [10],

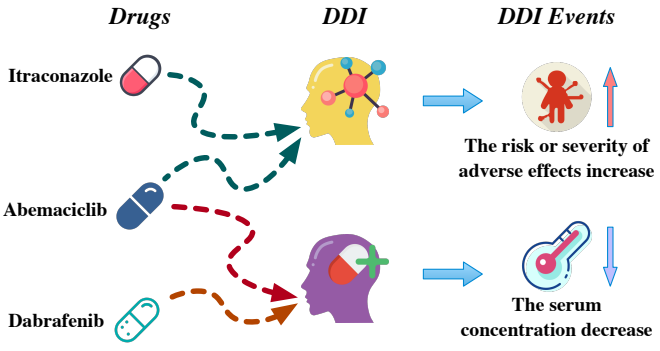

Fig. 1. An example of DDI event. When drug *Abemaciclib* and drug *Dabrafenib* interaction together, an DDI event will be occurred and cause the decrease of body's serum concentration. However, it will raise the risk or severity of adverse effects when mixing drug *Abemaciclib* and drug *Itraconazole*.

- *Tengfei Lyu, Jianliang Gao and Ling Tian are with the School of Computer Science and Engineering, Central South University, Changsha 410083, Hunan, China.*
  E-mail:{ tengfeilyu, gaojianliang, tianling}@csu.edu.cn
- *Zhao Li is with the Alibaba Group, Hangzhou 311121, Zhejiang, China.*
  E-mail: lizhao.lz@alibaba-inc.com
- *Peng Zhang is with the Institute of Cyberspace Advanced Technology, Guangzhou University, Guangzhou 510006, China.*
  E-mail: p.zhang@gzhu.edu.cn
- *Ji Zhang is with University of Southern Queensland, Australia.*
  E-mail: zhangji77@gmail.com

(Corresponding Author: Jianliang Gao.)

implementing multi-task learning on DDI type prediction [11], [12], [13], modeling semi-supervised learning to mine useful information for DDI prediction in both labeled and unlabeled drug data [7], and exploiting knowledge graph summarization for multi-typed DDI pharmacological effect prediction [14]. There have been also some efforts on predicting DDI using multiple data sources, such as the similarity features to obtain drug features for DDI events prediction task [5], [15], [16]. However, most existing methods pay less attention to the potential correlations between DDI events and other multimodal data such as targets and enzymes. Moreover, cross-modality complementarity of multimodal data has not been taken into consideration.

To incorporate multiple data sources, knowledge graphs are a powerful tool, and many biomedical knowledge bases have been published in this form [17]. Knowledge graphs have gained significant popularity due to successful applications to many different AI tasks such as link prediction [18], recommendation [19], healthcare misinformation [20], and bioinformatics [21]. Therefore, we have constructed a drug knowledge graph. In the drug knowledge graph, the edges have different semantic relations and the nodes represent different entities like drugs, targets, enzymes, substructures, transporters, and carriers. Drug knowledge graph, which is constructed from DrugBank can be used as an effective auxiliary for DDI events prediction, to find the inherent relations between drugs and related entities to improve prediction performance and provide explanations.

In this work, we aim to effectively assist the joint representation learning of multimodal data related to DDI events. We propose a **M**ultimodal **D**eep **N**eural **N**etwork (MDNN) framework for DDI events prediction. In MDNN, we design a two-pathway framework including drug knowledge graph (DKG) based pathway and heterogeneous feature (HF) based pathway to obtain drug multimodal representations. Then, inspired by graph neural network that try to learn from structure information [20], [22], [23], we propose the GNN layer to learn drug representations by extracting both structural information and semantic relations from the DKG. Finally, a multimodal fusion neural layer is designed to predict DDI events by exploring the complementary between the drug multimodal representations. Our contributions are summarized as follows:

- We constructed a drug knowledge graph, which can be used as an effective auxiliary for this task. Then, we propose a new multimodal deep neural network with a two-pathway framework including the drug knowledge graph based pathway and the heterogeneous feature based pathway. MDNN can predict DDI events by exploiting the associations between DDI events and multimodal representations.
- The MDNN framework mainly has the following merits: (a) MDNN learns the representations from multimodal data and mines the inter-modality similarities from multiple sources. (b) MDNN exploits the topological structure information and semantic relations with drug knowledge graph.
- We conduct extensive experiments on a real-world dataset to demonstrate the effectiveness of our model compared with classic and state-of-the-art methods.

The rest of the paper is organized as follows. The related work is discussed in Section 2. We formulate the problem in Section 3. Then, in Section 4, the MDNN model is introduced and explained in detail. After that, we demonstrate the experiments in Section 5. Finally we conclude our work and talk about future directions in Section 6.

2 Related Works

This section highlights related work in two aspects: drug-drug interaction prediction and graph neural network based methods in DDI prediction.

2.1 Drug-drug Interaction Prediction

DDI prediction is a fundamental task with applications in many areas such as clinical and pharmaceutical decisions [24], [25]. The research works which aim to improve DDI prediction can be summarized in two directions: integrating multiple drug features and applying deep learning techniques.

Thus, DDI prediction is important task in both clinical application and drug development. Many efforts have been taken on calculating the similarities by integrating multiple data sources and predicting DDI based on the fused similarity [26], [27]. Many computational methods have been proposed to predict potential DDIs based on the assumption that similar drugs tend to interact with other similar drugs. Vilar et al. present a new methodology applicable on a large scale that identifies novel DDIs based on molecular structural similarity to drugs involved in established DDIs [28]. Liu et al. developed a method based on the validated DDI data retrieved from DrugBank and extracted various types of features from each DDI [29]. In addition, the works [3], [30] integrate multiple drug features to calculate the similarities among drugs , and then predict DDI accurately based on the fused similarity. DINTO [31] is described a comprehensive ontology for DDI knowledge, which is the first formal representation of different types of DDIs and their mechanisms and its application in the prediction of DDIs. Li et al. propose a new systems pharmacology framework consisting of a new algorithm termed probability ensemble approach, through integrating the molecular chemical space, the pharmacological space, the gene annotations, in particular, the connectivity of biological networks, to predict effective drug combinations [32]. Through calculating the functional similarity from drug carriers, drug transporters, drug enzymes and drug targets, Ferdousi et al. developed an approach to discover new DDIs [33]. [5] proposes an integrative framework to fuse the similarities of drug features with proper weights and predict DDI. [15] proposes to learn accurate and interpretable similarity measurement from multiple types of drug features for DDI prediction. In addition, [16] proposes a framework DDIMDL that combines diverse drug features to build a model for predicting DDI events.

Recently, there has been growing interests in applying deep learning techniques for DDI prediction. Different from drug similarity obtained from multiple sources, a deep learning framework named DeepDDI [13] is proposed to use molecular structures of drugs as inputs for predicting DDI types. In the [34], they evaluate an approach to

potential DDIs prediction using link prediction methodology. The work [11] proposes a new multitask dyadic prediction model to predict adverse drug-drug interactions. MLRDA [7] develops a multi-task semi-supervised learning framework which effectively exploits information that is beneficial for DDI prediction in unlabeled drug data. However, they are limited in obtaining the rich features of drugs in structural information and semantic relations. Moreover, only a few methods take different drug features as independent data and do not take cross-modality complementarity into consideration. With comparison to the classic and deep learning methods, our proposed framework is able to automatically extract drug structural information and heterogeneous features from the multimodal data.

2.2 Graph Neural Network based Methods

Graph neural network (GNN) [35] which aim to extend the deep neural network to deal with arbitrary graph-structured data. Owing to the ability to learn more comprehensive node representations than the models that consider only node features [36] or graph structure [37], [38], graph neural network has been a promising results on various graph-based tasks in social networks [39], [40], text classification [41], [42], and recommendation [43], [44], [45] etc. GNN are powerful models for learning rich relational information in graphs by aggregating features from adjacent nodes/edges.

Inspired by the success of applying graph neural network in a wide variety of tasks [46], [47], [48], researchers also tried to utilize GNN to improve the performance of DDI events prediction. For example, Decagon [12] applies a relational GNN for predicting side effects of drug pairs. A semi-supervised graph autoencoder method is proposed by [15] that model the integration using multi-view graph autoencoders. Moreover, [49] is proposed with a novel neural method that extract drug-drug interactions from texts using external drug molecular structure information. Then, they utilize a graph convolutional network to encode the molecular structures. [50] integrates graph embedding methods for DDI prediction task. In addition, CASTER [10] develops an end-to-end dictionary learning framework for predicting DDI with chemical structures of drugs. Compared with traditional methods of integrating multiple data sources, the graph neural network model not only encodes the graph structure, but also encode the node features. Although these methods have achieved strong performance, a neglected disadvantage is that these model do not consider their multiple related correlations in the knowledge graph.

In addition, knowledge graph can provide a large amount of structured information among multiple entities and semantic relations associated with entities [51]. The core in the fusion of knowledge graph is how to effectively integrate such structural information into latent vector representations of drugs and related nodes. Knowledge graphs are a powerful tool [52], [53], and some biomedical knowledge bases have been published in this form. These knowledge graph-based methods also have been used in structured scenarios of DDI prediction. [54] have applied knowledge graph embedding approaches to extract feature vector representation of drugs to predict potential drug-drug interactions. KG-DDI [17] proposed with a new deep learning approach for predicting DDIs based on multiple data sources includding knowledge graph. KGNN [21] designs an effective framework for DDI prediction which can capture drug and its potential neighborhoods in the knowledge graph. Moreover, in SumGNN [14], they propose a knowledge summarization graph neural network method for multi-typed DDI predictions, which is mainly enabled by a local subgraph module that can efficiently anchor on relevant subgraphs from a knowledge graph.

Although these methods have achieved relatively good performance, they do not consider the drug multimodal data coherence and complementarity together. Compared with these methods, our model uses a newly designed graph neural network to capture both the topological information and semantic relations, and explores the cross-modality complementarity of multimodal data, which differentiates it from the existing methods.

3 Problem Formulation

In this section, we formulate the problem of DDI events prediction that we will tackle. We first present several basic definitions which will be used in the problem formulation.

DDI matrix. Formally, we denote DDI events $\mathcal{Y} \in (0, y_{ij})^{N_d \times N_d}$ as the label matrix for this prediction task, where N_d denotes the number of drugs in the DDI events matrix. $y_{ij} \in \mathcal{L}$ is a label, where $\mathcal{L} = \{y_1, y_2, \cdots, y_{N_l}\}$ denotes the label set and N_l denotes the types number of events. For each DDI event, $y_{ij} \in \mathcal{L}$ means that the interaction event y_{ij} exists between drug d_i and drug d_j, and $y_{ij} = 0$ means that there is no interaction event existing between drug d_i and drug d_j.

Drug Knowledge Graph (DKG). We consider a special type of knowledge graph for DDI events prediction named drug knowledge graph (DKG), denoted by $\mathcal{G} = (\mathcal{D}, \mathcal{R}, \mathcal{T})$:

$$\mathcal{G} = \{(d, r_{dt}, t) | d \in \mathcal{D}, r_{dt} \in \mathcal{R}, t \in \mathcal{T}, \mathcal{D} \cap \mathcal{T} = \varnothing\}, \quad (1)$$

where $\mathcal{D}$ and $\mathcal{T}$ describe a subset of drug entities and a subset of tail entities (drug related nodes, e.g. targets) respectively, and $\mathcal{R}$ denotes the set of relations between drugs and tail entities.

Heterogeneous Feature (HF). In this study, heterogeneous features consist of the target feature, substructure feature and enzyme feature. It is expressed as follows:

$$\mathcal{X}_d = \{X_t, X_s, X_e\} \in \mathbb{R}^{N_d \times (N_t + N_s + N_e)}, \quad (2)$$

where $X_t \in \mathbb{R}^{N_d \times N_t}$, $X_s \in \mathbb{R}^{N_d \times N_s}$ and $X_e \in \mathbb{R}^{N_d \times N_e}$ stand for the target feature matrix, the substructure feature matrix and the enzyme feature matrix, respectively. N_t, N_s and N_e represent the feature number of the targets, the substructures and the enzymes, respectively.

DDI Events Prediction. Given the DDI events matrix $\mathcal{Y}$, drug knowledge graph $\mathcal{G}$ and heterogeneous feature $\mathcal{X}_d$, we aim to predict specific interaction events between drug d_i and drug d_j. In other words, we formulate DDI events prediction as a multi-class classification problem. Our goal is to learn a prediction function $\hat{y}_{ij} = \Gamma(d_i, d_j | \Theta, \mathcal{Y}, \mathcal{G}, \mathcal{X}_d)$, where $\hat{y}_{ij}$ represents the probability of an event between drug d_i and drug d_j, and Θ denotes the model parameters of function Γ.

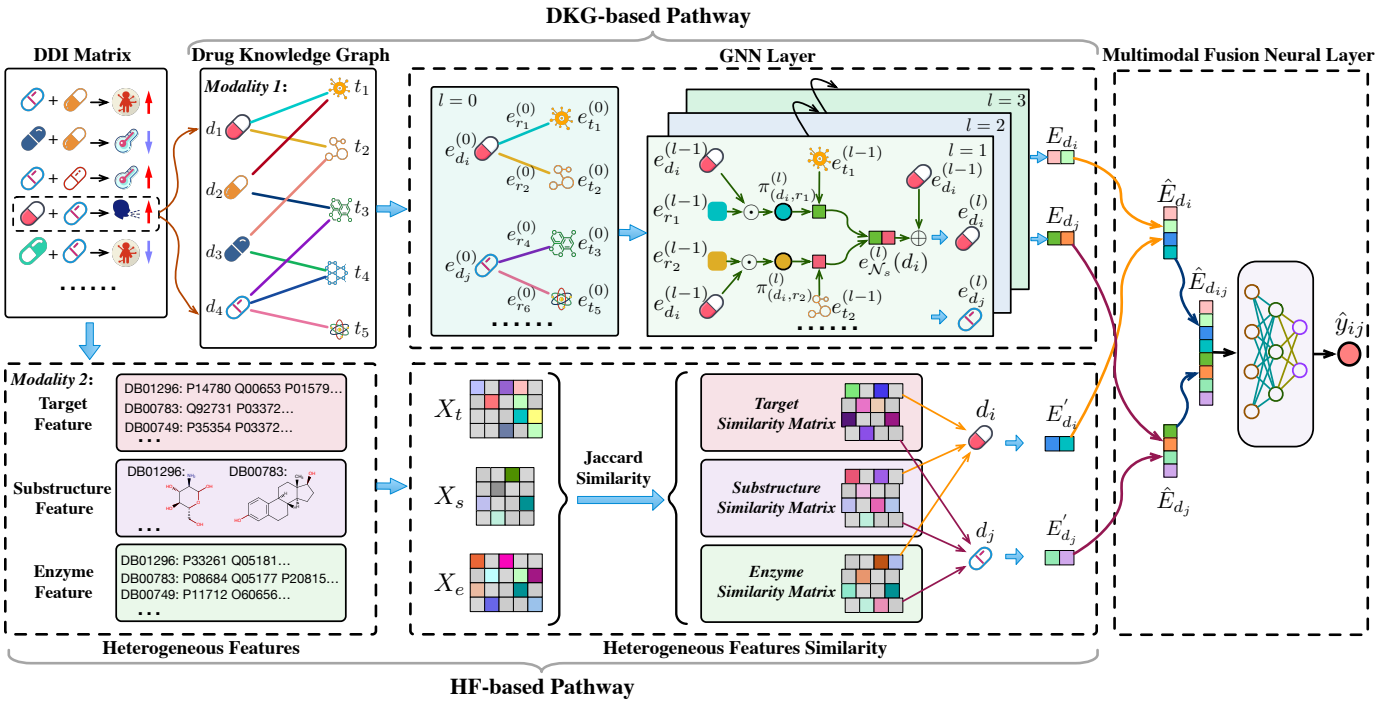

Fig. 2. Illustration of the proposed MDNN, consisting of two core pathways: the DKG-based pathway and the HF-based pathway. (1) The DKG-based pathway utilizes the graph neural network to extract the topological structural information and semantic relations from the constructed drug knowledge graph (DKG). (2) The HF-based pathway mines the inter-modality similarities of each heterogeneous feature from multiple sources. (3) The multimodal fusion neural layer is applied to effectively assist the joint representation learning of both the structural information and attribute feature, which explore the cross-modality complementarity of the multimodal data.

4 Methodology

Overview. The architecture of MDNN is depicted in Figure 2, which is composed of two main pathways: the DKG-based pathway and the HF-based pathway. The DKG-based pathway utilizes the graph neural network to extract the topological structure information and semantic relations between drugs on the constructed drug knowledge graph. The HF-based pathway aims to extract predictive information from different modalities to enhance the performance of the learned models. The multimodal fusion neural layer is applied to effectively assist the joint representation learning of both the structural information and the heterogeneous feature which explore the cross-modality complementarity of the multimodal data.

In the section, MDNN will be explained in details next, including the DKG-based pathway in subsection 4.1, the HF-based pathway in subsection 4.2 and the multimodal neural fusion layer in subsection 4.3.

4.1 The DKG-based Pathway

We explore the advantage of the abundant information related to the topological structure and semantic relations in the DKG, which is beneficial for DDI events prediction.

Drug Knowledge Graph. Since creating an integrated DKG is the indispensable step in our framework, we first focus on the drug knowledge graph (DKG) construction. For each drug in the DDI matrix, we collect the drug related entities on DrugBank, such as targets, transporters, etc. In order to obtain rich semantic information, we consider the general function of the tail entities as the relations between the drug and the tail entities. For example, the drug *DB05812* has a carrier named *serum albumin* (Uniprot ID: *P02768*), and the general function of *P02768* is *toxic substance binding*, leading to the triple of the DKG representation <*DB05812*, *toxic substance binding*, *P02768*>. In this way, we can obtain the drug knowledge graph triples (drug, relation, tail entity) with abundant information including the topological structure and semantic relations.

The GNN Layer. After constructing the knowledge graph, how to learn the topological structure information and the semantic information of the targeting relationship is the key step to obtain the embedded representation of the drug nodes. Taking into account the nonlinear modeling ability and automatic learning ability of graph neural network (GNN) technology, we decided to use GNN technology to perform feature mining and node representation learning on the drug knowledge graph. Therefore, in order to obtain a node vector representation with rich characterization information, the embedding representation of each drug node is iteratively updated by the attribute transfer of the neighbor nodes and the transfer of the semantic information of the edges in the drug knowledge graph. The GNN layer is proposed to capture drug topological structure and semantic relations in the drug knowledge graph. The initial representation matrix of the drug knowledge graph $\mathcal{G}$ is as follows:

$$E_{\mathcal{G}} = [\underbrace{e^{(0)}_{d_1}, \cdots, e^{(0)}_{N_d}}_{drug\ embedding}, \underbrace{e^{(0)}_{r_1}, \cdots, e^{(0)}_{N_r}}_{relation\ embedding}, \underbrace{e^{(0)}_{t_1}, \cdots, e^{(0)}_{N_k}}_{tail\ embedding}], \quad (3)$$

where N_d, N_r and N_k represent the number of drugs, relations and tail entities in the DKG, respectively. $e_d^{(0)} \in \mathbb{R}^d$, $e_r^{(0)} \in \mathbb{R}^d$ and $e_t^{(0)} \in \mathbb{R}^d$ are served as the initialization of drug embedding, relation embedding and tail entity embedding, respectively, where d is the dimension of embedding in drug knowledge graph.

Consider a candidate drug pair of drug d_i and drug d_j. We are aimed at extracting information underlying the graph-structured data into low dimensional vector representations. We can also further stack more GNN layers to explore the topological structure information, gathering the information propagated from the higher-hop neighbors. To consider efficiency and the fixed computation pattern of each batch data, we uniformly sample a fixed size set $\mathcal{N}_s(d_i)$ instead of using its full neighbors.

For each drug d_i, we uniformly sample a set of a fixed size as $\mathcal{N}_s(d_i)$ instead of using all the neighbors. It is important to explicitly incorporate the semantics of relations into drug representation learning. Thus, we compute the semantics feature score between drug d_i and tail entity t_n with relation r_{in} as follow:

$$\pi^{(l)}_{(d_i,r_{in})} = sum[(e^{(l-1)}_{d_i} \odot e^{(l-1)}_{r_{in}})W_1^{(p)} + b_1^{(p)}] \tag{4}$$

where $e^{(l-1)}_{r_{in}}$ is the relation representation between drug d_i and tail entity t_n after $(l-1)^{th}$ GNN layer. $e^{(l-1)}_{d_i}$ is the drug d_i representation generated from the previous message-passing steps, memorizing the messages from its (l-1)-hop neighbors. $W_1^{(p)}$ is the trainable weight matrix, $b_1^{(p)}$ is the bias vector and p is the number of full connection layers, $\odot$ denotes the element-wise product.

Then, we aggregate the messages propagated from the neighborhood $\mathcal{N}_s(d_i)$ to refine the embedding of d_i. More formally, we first recursively formulate the neighborhood representation of drug d_i at l^{th} layer. We define the neighborhood aggregation function as:

$$e^{(l)}_{\mathcal{N}_s(d_i)} = \sum_{t_n \in \mathcal{N}_s(d_i)} \pi^{(l)}_{(d_i,r_{in})} e^{(l-1)}_{t_n} \tag{5}$$

The final step aggregates the embedding of drug $e^{(l-1)}_{d_i}$ and its neighborhood embedding $e^{(l)}_{\mathcal{N}_s(d_i)}$ into a vector using the following aggregation function:

$$E_{d_i} = e^{(l)}_{d_i} = \sigma((e^{(l-1)}_{d_i} \oplus e^{(l)}_{\mathcal{N}_s(d_i)})W_2 + b_2), \tag{6}$$

where $W_2 \in \mathbb{R}^{(2d)\times d}$ is the trainable weight matrix and σ is the activation function $ReLU$. $\oplus$ denotes the concatenate operation.

Similarly, we can obtain the representation E_{d_j} for drug d_j by propagating information from its neighboring nodes. In summary, the advantage of the embedding propagation layer lies in explicitly exploiting the first-order connectivity information for drug representations.

4.2 The HF-based Pathway

In the HF-based pathway, we use heterogeneous features to calculate the drug similarity between DDI events. Each feature corresponds to a set of descriptors, and thus a drug can be represented by a binary feature vector, whose each entry (1 or 0) indicates the presence or absence of the corresponding descriptor. In order to make the drug node representation more dense and improve the accuracy of the vector, we use principal components analysis (PCA) to compress features and reduce the sparsity. We calculate the pairwise drug–drug similarity from feature vectors using the Jaccard similarity measurement.

$$J(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} = \frac{|d_i \cap d_j|}{|d_i| + |d_j| - |d_i \cap d_j|} \tag{7}$$

By using the Jaccard similarity measurement, we can obtain the target similarity matrix $E^t \in \mathbb{R}^{N_d \times k}$, substructure similarity matrix $E^s \in \mathbb{R}^{N_d \times k}$ and enzyme similarity matrix $E^e \in \mathbb{R}^{N_d \times k}$, where N_d stands for the number of drugs, and the superscript k denotes the dimension of heterogeneous feature embedding.

After obtaining the similarity matrix, we can get the embedding of drug d_i as $e^t_{d_i} \in E^t$, $e^s_{d_i} \in E^s$ and $e^e_{d_i} \in E^e$, respectively.

Finally, to further explore the inter-modal complementarity of the heterogeneous features, we concatenate the three representation vectors as the final heterogeneous features embedding of d_i, which is formulated as:

$$E^{'}_{d_i} = e^t_{d_i} \oplus e^s_{d_i} \oplus e^e_{d_i} \tag{8}$$

Similarly, the embedding $E^{'}_{d_j}$ of drug d_j can be obtained.

Algorithm 1: MDNN algorithm

Input: DDI matrix $\mathcal{Y}$; drug knowledge graph $\mathcal{G}$; heterogeneous features $\mathcal{X}_d$; hyper-parameter: l, $\mathcal{N}_s$, d and k.
Output: $\Gamma(d_i, d_j|\Theta, \mathcal{Y}, \mathcal{G}, \mathcal{X}_d)$

```
Initialization G
while MDNN not converge do
  for (d_i, d_j) in Y do
    // DKG-based pathway
    N_0 ← e, ∀e ∈ D ∪ T
    {N_l}_{l=1}^L ← Neighborhood Sampling (entity e)
    e^(0) ← e, ∀e ∈ N_0
    for l = 1, ··· , L do
      for e ∈ N_l do
        e^(l)_{N_l} ← Σ_{t_n ∈ N_l(e)} π^(l)_(e,r_in) e^(l-1)_{t_n}
        e^(l) ← e^(l-1) ⊕ e^(l)_{N_l}
      end for
    end for
    E_{d_i} ← e^(l)_{d_i}, E_{d_j} ← e^(l)_{d_j}
    // HF-based pathway
    for X ∈ X_d do
      e^x ← Jaccard Similarity (X)
    end for
    E'_{d_i} ← e^t ⊕ e^s ⊕ e^e, E'_{d_j} ← e^t ⊕ e^s ⊕ e^e
    Ê_{d_i} ← E_{d_i} ⊕ E'_{d_i}, Ê_{d_j} ← E_{d_j} ⊕ E'_{d_j}
    Calculate ŷ_ij = f(Ê_{d_i} ⊕ Ê_{d_j})
    Update parameters Θ
  end for
end while
return Γ
```

4.3 Multimodal Neural Fusion Layer

Intuitively, the DKG-based and the HF-based pathways provide complementary information to each other. To achieve the best utilization of the information of these two pathways, we consider their coherence and complementarity together in the so-called multimodal neural fusion layer. After obtaining the embedding E_{d_i} and $E^{'}_{d_i}$ for drug d_i, these embedding are linked together as the final multimodal embedding $\hat{E}_{d_i}$ of drug d_i. The equation can be described as:

$$\hat{E}_{d_i} = E_{d_i} \oplus E^{'}_{d_i}. \tag{9}$$

As such, the embedding of d_i contains not only the heterogeneous feature information, but also the semantic information of the relation and its structural information. Similarly, the final embedding $\hat{E}_{d_j}$ of drug d_j can be obtained.

Then, the multimodal fusion embedding $\hat{E}_{d_{ij}}$ with multiple fully connected layers is used to predict the DDI events:

$$\hat{y}_{ij} = \rho((\hat{E}_{d_i} \oplus \hat{E}_{d_j})W_3^{(q)} + b_3^{(q)}), \tag{10}$$

where $W_3^{(q)}$ is the trainable weight matrix, and $b_3^{(q)}$ is the bias vector, q is the number of the full connected layers. ρ denotes the activation function $softmax$. Finally, we use softmax function and obtain the final prediction score $\hat{y}_{ij}$. To summarize, the training procedure of our MDNN is presented in Algorithm 1.

For model optimization, we add batch normalization layers to accelerate the convergence, and add dropout layers to avoid over-fitting and enhance generalization ability. And we adopt cross-entropy as the loss function, and empirically train and optimize the MDNN model. In addition, we use L2 regularization to prevent over-fitting of our model.

5 Experiments

In this section, we first describe the experimental dataset, baselines, evaluation metrics, parameter and evaluation settings. Then, we show that MDNN consistently outperforms the state-of-the-art baselines in terms of both prediction accuracy and explainability. Ablation study and parameter sensitivity analysis are also conducted to demonstrate how the components in the framework contribute to the performance. We also investigate the effects of our proposed model using different task and present a multi-task analysis with our method.

5.1 Experimental Setup

Dataset. In order to demonstrate the effectiveness of our proposed model on DDI events prediction, we conduct extensive experiments on a real-world dataset including three parts: DDI matrix, drug knowledge graph and heterogeneous features. Table 1 shows the main statistics of the dataset. We describe their properties below:

- **DDI Matrix:** We obtain the verified DDI events data from DDIMDL [1], which contains 572 drugs and 65 types of events. According to statistics, a total of 37,264 drug pairs have definite drug-drug interactions.
- **Drug Knowledge Graph:** According to DDI events, we collect the drug knowledge graph from the DrugBank[2] (version 5.1.7). It is a real-world dataset that contains 572 drugs and 1,614 entities with 76,871 triplets and 157 relations between drugs and tail nodes.
- **Heterogeneous Features:** Heterogeneous features released by DDIMDL [16] include 1,162 target features, 583 substructure features and 202 enzyme features.

TABLE 1
The Overall Information of Dataset

Category	Name	Number
DDI Matrix	Drugs	572
	Events	65
	Drug-Drug Pairs	37,264
Drug Knowledge Graph	Triples	76,871
	Drugs	572
	Relations	157
	Tail Nodes	1042
Heterogeneous Features	Targets	1,162
	Substructures	583
	Enzymes	202

Baselines. In this study, we compare our model against the following baselines, including the traditional and the recent state-of-the-art methods. The baselines are as follows:

- **DDIMDL**: DDIMDL [16] adopts a joint deep neural network framework to learn the representations of drug–drug pairs and predict DDI events.
- **KGNN**: KGNN [21] samples and aggregates neighborhoods for each node from their local receptives via GNN and with external KG, which achieves the state-of-the-art result on binary DDI prediction problem.
- **DeepDDI**: DeepDDI [13] develops a deep learning-based method that reduces the dimension of drug features based on a principal component analysis.
- **Other traditional methods**: In this study, we consider several traditional classification approaches, i.e., random forest, k-nearest neighbour, logistic regression and deep neural network.

Evaluation Metrics. We evaluate the prediction performance using several multi-class classification evaluation metrics, including accuracy (Acc), area under the precision–recall-curve (AUPR), area under the ROC curve (AUC), F1 score (F1), Precision (Pre) and Recall (Rec). We use micro metrics for AUPR and AUC, while macro metrics for F1 score, Precision and Recall. Notion that, micro-Precision, micro-Recall and micro-F1 are equal to accuracy in the multi-class problem.

Parameter and Evaluation Settings. The maximum iteration number is set to 100, and we use a batch size of 1,024 and adopt Adam algorithm with a learning rate of 0.001 to optimize all trainable parameters through a random search in each iteration. We set the neighborhood sampling

1. https://github.com/YifanDengWHU/DDIMDL

2. https://go.drugbank.com/

TABLE 2
Performance of MDNN Against Comparative Approaches

Methods	Acc	AUPR	AUC	F1	Pre	Rec
Logistic Regression	0.7920	0.8400	0.9960	0.5948	0.7437	0.5236
K-Nearest Neighbour	0.7214	0.7716	0.9813	0.4831	0.7174	0.4081
Random Forest	0.7775	0.8349	0.9956	0.5936	0.7893	0.5161
Deep Neural Network	0.8797	0.9134	0.9963	0.7223	0.8047	0.7027
DeepDDI	0.8371	0.8899	0.9961	0.6848	0.7275	0.6611
KGNN	0.8782	0.9357	0.9969	0.7443	0.7845	0.7284
DDIMDL	0.8852	0.9208	0.9976	0.7585	0.8471	0.7182
MDNN	**0.9175**	**0.9668**	**0.9984**	**0.8301**	**0.8622**	**0.8202**

$\mathcal{N}_s = 6$, GNN layers $l = 1$, L2 weight = 1e-8, $p = 2$, $q = 3$, dimension $d = 128$ and $k = 256$.

To comprehensively evaluate our proposed method, we adopt 5-fold cross validation and randomly divide all DDI pairs into five subsets in our experiments. The evaluation score is the average of the output of the five rounds. We use the early-stopping strategy to prevent over-fitting which automatically stops the training if no improvement is observed after 10 epochs.

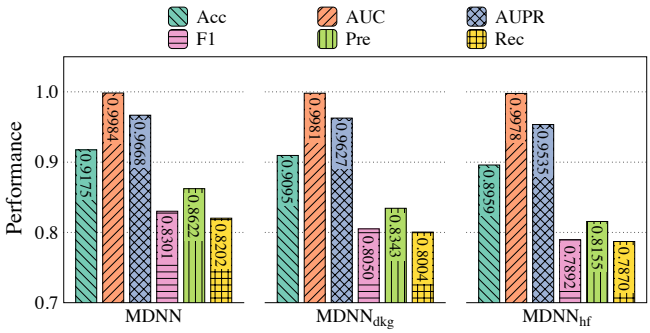

Fig. 3. Results of the ablation experiments with the relative performances compared with complete MDNN in all the metrics. MDNN$_{dkg}$ and MDNN$_{hf}$ mean that only using DKG-based and HF-based pathway to make prediction, respectively.

5.2 Results and Analysis

In this section, we report the performance of our model and all baselines in Table 2. From the result, we find that our model achieves the best performance in DDI events prediction on the real-world dataset. Compared to all the baselines, our MDNN model achieves significant improvements in terms of all the evaluation metrics. Particularly, our proposed model outperforms DDIMDL by 3.23% on Acc, 4.6% on AUPR, 0.08% on AUC, 7.16% on F1, 1.51% on Pre and 10.2% on Rec. The obvious promotion on the DDIMDL model confirms that it is not only the topological structural information but also semantic relation that produce accurate drug embeddings. Furthermore, the multimodal information fusion is also important. In addition, comparing with the baseline KGNN, our MDNN model still performs much better. our model improves KGNN by 3.93% on Acc, 3.11% on AUPR, 0.15% on AUC, 8.85% on F1, 7.77% on Pre and 9.18% on Rec. This demonstrates the high effectiveness of MDNN. These results prove that the MDNN model proposed in two pathway is also effective for DDI events prediction and achieves the best performance.

The better performance of our model is attributed to the fact that our model explores both the drug topological embedding representations in the drug knowledge graph and the cross-modality embedding representations of the multimodal data. Moreover, the comparative study with other state-of-art methods demonstrates that our model achieves the most stable performance which may be due to (a) MDNN incorporates a GNN model to exploit the topological structure information and semantic relations in the drug knowledge graph; (b) MDNN leverages the cross-modality complementary information of the multimodal data. To sum up, it is a preferable achievement in terms of DDI events prediction.

5.3 Ablation Study

The MDNN model includes two main components: the DKG-based pathway and HF-based pathway. To explore how the DKG-based and the HF-based pathways improve the performance of the proposed model, we conduct the ablation study on the following variants of MDNN. Two variants of MDNN are designed to further explore the different contribution components of MDNN. The description of these two variants is as follows: (1) MDNN$_{dkg}$ is the model variant where we only consider the topological structures and semantic relations to learn the embedding of drug–drug pairs from the DKG. (2) MDNN$_{hf}$ is the model variant where we only explore cross-modality embedding of drug–drug pairs using only the heterogeneous features including target, sunstructure and enzyme features. We illustrate the experimental results in Figure 3 and make some discussions.

From Figure 3, we observe that MDNN outperforms its variants, which separately disregard different components of MDNN. MDNN$_{dkg}$ and MDNN$_{hf}$ are two components for DDI events prediction of MDNN, they ignore the cross-modality embedding fusion of multimodal features and the topological structural information from drug knowlwdge graph, respectively. MDNN outperforms them both. This observation suggests the significance of integrating the multimodal features and the inter-modality representation from multiple sources in MDNN. Figure 3 shows that the ablation results which verify the contribution of each pathway in our model, showing that combining the topological representations in neighborhood with the semantic relations from

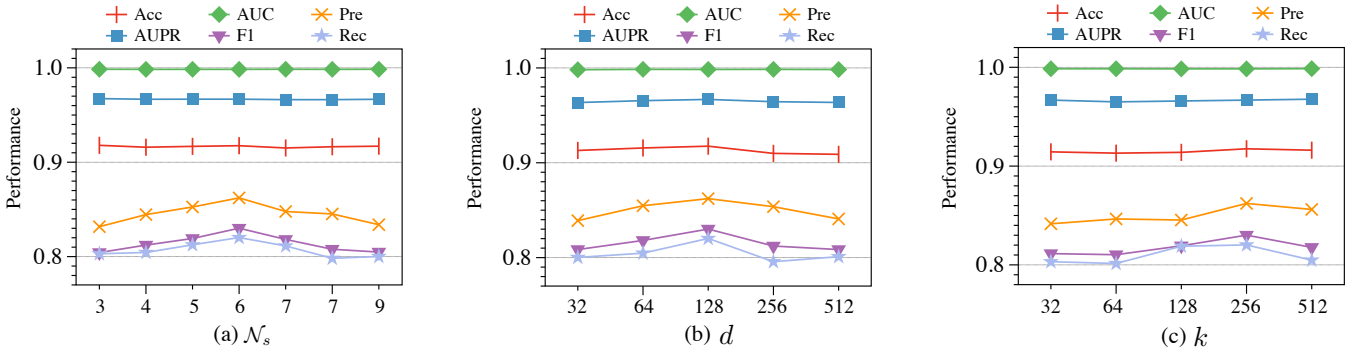

Fig. 4. Parameter sensitivity analysis on effect of neighborhood sample $\mathcal{N}_s$, initialization dimension d in DKG-based pathway and embedding dimension k in HF-based pathway.

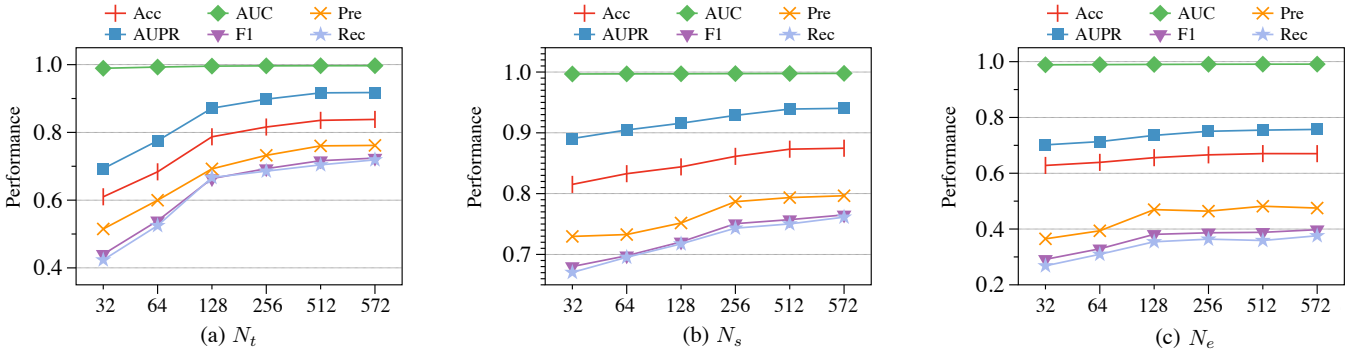

Fig. 5. Parameter sensitivity analysis on effect of each heterogeneous feature. N_t, N_s and N_e indicate that only use targets feature, substructures feature and enzymes feature to explore the effect of different embedding dimensions respectively.

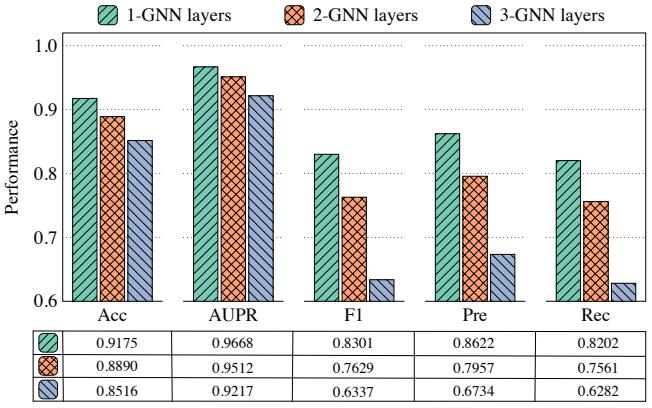

Fig. 6. Parameter sensitivity analysis of different number of GNN layers in the DKG-based pathway.

the DKG and the heterogeneous features is beneficial to improving the DDI events prediction performance.

5.4 Parameter Sensitivity Analysis

In this work, there are four essential parameters, which are the size of neighborhood sample $\mathcal{N}_s$, the number of GNN layers l, the dimension of initialization embedding d in DKG-based pathway and the dimension of embedding k in HF-based pathway. We fixed other parameters when studying the effect of each of them. The results are shown in Figure 4 and Figure 6. In addition, we explore the effect of each heterogeneous feature when only using target feature, substructure feature and enzyme feature respectively. The results are shown in Figure 5.

Effect of GNN layers. We investigate the influence of the GNN layer l by varying its value from 1 to 3. From Figure 6, we observe that the performance of our model in all the metrics decreases starting from l = 1, as a larger l brings massive noises to the model. This is also in line with our intuition that using nodes with too many hops makes little difference when encoding the topological information of each drug in DKG. The experiment results implies that $l = 1$ is often ideal for real cases.

Effect of neighborhood size. We vary the size of sampled neighbor $\mathcal{N}_s$ to explore the efficacy of MDNN. Figure 4(a) shows that our model achieves the best performance when $\mathcal{N}_s$ = 6. When $\mathcal{N}_s$ is too small, the model cannot fully incorporate the structural information, while an large value of $\mathcal{N}_s$ makes the model more prone to be misled by noises.

Effect of initialization dimension in DKG-based pathway. In addition, we examine the influence of embedding dimension d in the drug knowledge graph by varying from its value from 32 to 512. The performance for the prediction achieves peak value when d = 128, shown in Figure 4(b). Intuitively, the performance can be enhanced with the proper d that can encode enough information of drugs and entities from the DKG. When d is too large, however, the model will be affected by the over-fitting.

Effect of embedding dimension in HF-based pathway. To study the impact of embedding dimension in HF-based

pathway, we investigate how the proposed framework works with the changes of embedding dimension, while fixing other parameters. Figure 4(c) shows the sensitivity of embedding dimension in HF-based pathway. First, the performance for the DDI events prediction achieves peak when k = 256. This indicates that heterogeneous features indeed benefit the performance, but too large value of heterogeneous features embedding dimension may cause the noises to increase.

Effect of each heterogeneous feature. In this experiment, we evaluated the influence of different drug heterogeneous features on the DDI event prediction. our goal is to verify the effectiveness of each heterogeneous feature in different dimensions. In other words, it means that we only consider different dimensions of one heterogeneous feature for this task. We first use target features, substructure features and enzyme features of drug to calculate the Jaccard similarity matrix, respectively. Then we can obtain embedding representations of different dimensions after using PCA to reduce the dimensionality. In the Figure 5, N_t, N_s and N_e represent that we examine the influence of embedding dimension by varying from its value from 32 to 572 only using target features, substructure features and emzyme features, respectively. Intuitively, when one of heterogeneous features is used for prediction, the larger the dimension of the embedded representation, the more effective information it contains, and the better the prediction effect of the model. However, the model will be misled by noises when using target, substructure and enzyme features at the same time for DDI events prediction.

TABLE 3
Performance Comparison of MDNN with Other Methods on Two Different Tasks

Task	Methods	Acc	AUPR	F1	Rec
Task A	DNN	0.6239	0.6361	0.2997	0.2840
	DeepDDI	0.5774	0.5594	0.3416	0.3890
	DDIMDL	0.6415	0.6558	0.4460	0.4319
	MDNN	**0.6495**	**0.6661**	**0.4471**	**0.4611**
Task B	DNN	0.4087	0.3776	0.1152	0.1093
	DeepDDI	0.3602	0.2781	0.1373	0.1450
	DDIMDL	0.4075	0.3635	0.1590	0.1452
	MDNN	**0.4575**	**0.4215**	**0.1697**	**0.1709**

5.5 Multi-task Analysis

In this subsection, we further analyze the effectiveness of our proposed framework in other task. We created two different tasks by randomly splitting the drugs involved into five subsets and using four of them as the training drug set while the remaining one as the test drug set to evaluate the effectiveness of our model. For task A, prediction models are constructed on the DDIs between training drugs, and then make predictions for DDIs between training drugs and test drugs. For task B, prediction models are constructed on the DDIs between training drugs, and then make predictions for DDIs between test drugs. Task A is suitable for the prediction of the interaction between newly developed drugs and the drugs with known pharmacological side effects, while task B is suitable for the prediction of interactions between newly developed drugs.

It can be learned from Table 3 that the experimental results of our model in both tasks are better than other methods. This effectively shows that whether it is between known drugs or new drugs, the utilization of structural information and heterogeneous features improves the prediction accuracy of DDI events, and provides a strong, reliable support for research on DDI events prediction.

6 Conclusion and Discusion

In this paper, we propose a new MDNN model including two mainly pathway for drug-drug interaction events prediction. In the DKG-based pathway, to make the most usage of the abundant structural information, we firstly construct a drug knowledge graph. Then, the GNN layer uses a graph neural network to learn drug and its topological neighborhood representations from the DKG, which extracts both structural information and semantic relations. In the HF-based pathway, we aim to extract predictive information from different modalities to enhance the performance of the learned models. MDNN effectively exploits both the topological information and the semantic relations by leveraging a graph neural network on the drug knowledge graph. Moreover, MDNN also exploits the joint representation learning of both the structure information and the heterogeneous features, which effectively explores the cross-modal complementarity of the multimodal data. The experimental results show that MDNN outperforms the classic and state-of-the-art DDI events prediction models.

Our future work on this problem will mainly focus on further enhancing the prediction accuracy of DDIs and developing more efficient solutions that can support larger drug databases to fully exploit of the power of deep learning without overfitting. The current work essentially establishes a solid, extendable foundation for us to pursue these future directions.

References

[1] K. M. Giacomini, R. M. Krauss, D. M. Roden, M. Eichelbaum, M. R. Hayden, and Y. Nakamura, "When good drugs go bad," *Nature*, vol. 446, no. 7139, pp. 975–977, 2007.

[2] T. Zhang, J. Leng, and Y. Liu, "Deep learning for drug–drug interaction extraction from the literature: a review," *Briefings in bioinformatics*, vol. 21, no. 5, pp. 1609–1627, 2020.

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

[4] S. Vilar, C. Friedman, and G. Hripcsak, "Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media," *Briefings in bioinformatics*, vol. 19, no. 5, pp. 863–877, 2018.

[5] P. Zhang, F. Wang, J. Hu, and R. Sorrentino, "Label propagation prediction of drug-drug interactions based on clinical side effects," *Scientific reports*, vol. 5, no. 1, pp. 1–10, 2015.

[6] M. Moridi, M. Ghadirinia, A. Sharifi-Zarchi, and F. Zare-Mirakabad, "The assessment of efficient representation of drug features using deep learning for drug repositioning," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–11, 2019.

[7] X. Chu, Y. Lin, Y. Wang, L. Wang, J. Wang, and J. Gao, "Mlrda: A multi-task semi-supervised learning framework for drug-drug interaction prediction," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2019, pp. 4518–4524.

[8] X. Liang, P. Zhang, J. Li, Y. Fu, L. Qu, Y. Chen, and Z. Chen, "Learning important features from multi-view data to predict drug side effects," *Journal of Cheminformatics*, vol. 11, no. 1, pp. 1–17, 2019.

[9] G. Zahoránszky-Kőhalmi, T. Sheils, and T. I. Oprea, "Smartgraph: a network pharmacology investigation platform," *Journal of Cheminformatics*, vol. 12, no. 1, p. 5, 2020.

[10] K. Huang, C. Xiao, T. Hoang, L. Glass, and J. Sun, "Caster: Predicting drug interactions with chemical substructure representation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 702–709.

[11] B. Jin, H. Yang, C. Xiao, P. Zhang, X. Wei, and F. Wang, "Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017, pp. 1367–1373.

[12] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.

[13] J. Y. Ryu, H. U. Kim, and S. Y. Lee, "Deep learning improves prediction of drug-drug and drug-food interactions," *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, pp. E4304–E4311, 2018.

[14] Y. Yu, K. Huang, C. Zhang, L. M. Glass, J. Sun, and C. Xiao, "Sumgnn: Multi-typed drug interaction prediction via efficient knowledge graph summarization," *Bioinformatics*, pp. 1–8, 2021.

[15] T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive multi-view graph auto-encoders," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2018, pp. 3477–3483.

[16] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, "A multimodal deep learning framework for predicting drug–drug interaction events," *Bioinformatics*, vol. 36, no. 15, pp. 4316–4322, 2020.

[17] M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. 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*, 2019, pp. 113–123.

[18] J. Chen, H. He, F. Wu, and J. Wang, "Topology-aware correlations between relations for inductive link prediction in knowledge graphs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, 2021, pp. 6271–6278.

[19] S. Zhou, X. Dai, H. Chen, W. Zhang, K. Ren, R. Tang, X. He, and Y. Yu, "Interactive recommender system via knowledge graph-enhanced reinforcement learning," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 179–188.

[20] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 492–502.

[21] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "Kgnn: Knowledge graph neural network for drug-drug interaction prediction," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2020, pp. 2739–2745.

[22] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in neural information processing systems*, 2017, pp. 1024–1034.

[23] H. Wang, M. Zhao, X. Xie, W. Li, and M. Guo, "Knowledge graph convolutional networks for recommender systems," in *Proceedings of the World Wide Web Conference*, 2019, pp. 3307–3313.

[24] Q. Luo, S. Mo, Y. Xue, X. Zhang, Y. Gu, L. Wu, J. Zhang, L. Sun, M. Liu, and Y. Hu, "Novel deep learning-based transcriptome data analysis for drug-drug interaction prediction with an application in diabetes," *BMC bioinformatics*, vol. 22, no. 1, pp. 1–15, 2021.

[25] A. Nader, N. M. Mostafa, F. Ali, and M. Shebley, "Drug–drug interaction studies of elagolix with oral and transdermal low-dose hormonal add-back therapy," *Clinical Pharmacokinetics*, vol. 60, no. 1, pp. 133–143, 2021.

[26] E. Callegari, J. Lin, S. Tse, T. C. Goosen, and V. Sahasrabudhe, "Physiologically-based pharmacokinetic modeling of the drug–drug interaction of the ugt substrate ertugliflozin following co-administration with the ugt inhibitor mefenamic acid," *CPT: pharmacometrics & systems pharmacology*, vol. 10, no. 2, pp. 127–136, 2021.

[27] H. Wu, Y. Xing, W. Ge, X. Liu, J. Zou, C. Zhou, and J. Liao, "Drug-drug interaction extraction via hybrid neural networks on biomedical literature," *Journal of biomedical informatics*, vol. 106, p. 103432, 2020.

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

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

[30] I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, and M. Sadoghi, "Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions," *Journal of Web Semantics*, vol. 44, pp. 104–117, 2017.

[31] M. Herrero-Zazo, I. Segura-Bedmar, J. Hastings, and P. Martinez, "Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms," *Journal of chemical information and modeling*, vol. 55, no. 8, pp. 1698–1707, 2015.

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

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

[34] A. Kastrin, P. Ferk, and B. Leskošek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning," *PloS one*, vol. 13, no. 5, pp. 1–23, 2018.

[35] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[36] C. Yang, J. Liu, and C. Shi, "Extract the knowledge of graph neural networks and go beyond it: An effective knowledge distillation framework," in *Proceedings of the Web Conference 2021*, 2021, pp. 1227–1237.

[37] X. Ma, G. Qin, Z. Qiu, M. Zheng, and Z. Wang, "Riwalk: fast structural node embedding via role identification," in *2019 IEEE International Conference on Data Mining*, 2019, pp. 478–487.

[38] T. Fu, C. Xiao, X. Li, L. M. Glass, and J. Sun, "Mimosa: Multi-constraint molecule sampling for molecule optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 125–133.

[39] C. Huang, H. Xu, Y. Xu, P. Dai, L. Xia, M. Lu, L. Bo, H. Xing, X. Lai, and Y. Ye, "Knowledge-aware coupled graph neural network for social recommendation," in *35th AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4115–4122.

[40] H. Ji, J. Zhu, X. Wang, C. Shi, B. Wang, X. Tan, Y. Li, and S. He, "Who you would like to share with? a study of share recommendation in social e-commerce," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 232–239.

[41] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2016, pp. 2873–2879.

[42] P. Liu, X. Qiu, and X.-J. Huang, "Adversarial multi-task learning for text classification," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, 2017, pp. 1–10.

[43] Z. Wang, G. Lin, H. Tan, Q. Chen, and X. Liu, "Ckan: Collaborative knowledge-aware attentive network for recommender systems," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 219–228.

[44] B. Jin, C. Gao, X. He, D. Jin, and Y. Li, "Multi-behavior recommendation with graph convolutional networks," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 659–668.

[45] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 639–648.

[46] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, and Y. Zhao, "Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2020, pp. 1324–1330.

[47] Y. Song, S. Zheng, Z. Niu, Z.-H. Fu, Y. Lu, and Y. Yang, "Communicative representation learning on attributed molecular graphs," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2020, pp. 2831–2838.
[48] Z. Hao, C. Lu, Z. Huang, H. Wang, Z. Hu, Q. Liu, E. Chen, and C. Lee, "Asgn: An active semi-supervised graph neural network for molecular property prediction," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 731–752.
[49] M. Asada, M. Miwa, and Y. Sasaki, "Enhancing drug-drug interaction extraction from texts by molecular structure information," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 2, 2018, pp. 680–685.
[50] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S. M. Lin, W. Zhang, P. Zhang, and H. Sun, "Graph embedding on biomedical networks: methods, applications and evaluations," *Bioinformatics*, vol. 36, no. 4, pp. 1241–1251, 2020.
[51] D. N. Nicholson and C. S. Greene, "Constructing knowledge graphs and their biomedical applications," *Computational and structural biotechnology journal*, vol. 18, p. 1414, 2020.
[52] Y. Zhao, J. Zhang, Y. Zhou, and C. Zong, "Knowledge graphs enhanced neural machine translation," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2020, pp. 4039–4045.
[53] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "Kgat: Knowledge graph attention network for recommendation," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 950–958.
[54] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019.

Tengfei Lyu is a graduate student of the School of Computer Science and Engineering, Central South University, Changsha, China. He received the B.S. degree from Bohai University, Jinzhou, Liaoning, China, in 2019. His research interests are bioinformatics, machine learning and graph neural network.

Jianliang Gao received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, China. He is currently a professor with the School of Computer Science and Engineering, Central South University, China. His main research interests include bioinformatics, machine learning and graph data mining.

Ling Tian received the MS degree in the School of Computer Science and Engineering, Central South University, Changsha, China. Her main research focuses on bioinformatics and graph embedding.

Zhao Li received the Ph.D. degree (Hons.) from the Computer Science Department, University of Vermont. He is currently a Senior Staff Scientist with the Alibaba Group, specializing in ecommerce ranking and recommendation systems. He has published several articles in prestigious conferences and journals, including NIPS, AAAI, IJCAI, SIGIR and TKDE.

Peng Zhang is a professor with Guangzhou University. He received his PhD degree from University of the Chinese Academy of Sciences. He was a senior staff engineer with the Alibaba Group, a lecturer with University of Technology Sydney, and an associate professor with the Chinese Academy of Sciences. He has been researching into data mining, data streams, and social network analysis, with over 150 publications in TPAMI, TKDE, TNNLS, KDD, SIGIR, WWW, ICDM, AAAI, IJCAI, etc. He has served on many program committees of international conferences, including PC member for KDD, ICLR, ICML, NeurIPS, IJCAI, and AAAI conferences. He also served as the founding editorial board of Springer Annals of Data Science, and Springer Journal of Big Data.

Ji Zhang is currently a full Professor in computer science at the University of Southern Queensland (USQ), Australia. He is an IET Fellow, IEEE Senior Member, ACM member, Australian Endeavour Fellow, Queensland International Fellow (Australia) and Izaak Walton Killam Scholar (Canada). He is also an adjunct senior researcher of Zhejiang Lab, China. His research interests are big data analytics, knowledge discovery and data mining (KDD), computational intelligence, information privacy and security. He received his degree of Ph.D. from the Faculty of Computer Science at Dalhousie University, Canada in 2009. He has published over 200 papers in major peer-reviewed international journals and conferences.