MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events

Tengfei Lyu[1], **Jianliang Gao**[1*], **Ling Tian**[1], **Zhao Li**[2], **Peng Zhang**[3] and **Ji Zhang**[4]
[1]School of Computer Science and Engineering, Central South University
[2]Alibaba Group, Hangzhou, China
[3]Cyberspace Institute of Advanced Technology, Guangzhou University
[4]Zhejiang Lab, Hangzhou, China
{tengfeilyu, gaojianliang, tianling}@csu.edu.cn, lizhao.lz@alibaba-inc.com,
zhangpeng04@gmail.com, zhangji77@gmail.com

Abstract

The interaction of multiple drugs could lead to serious events, which causes injuries and huge medical costs. Accurate prediction of drug-drug interaction (DDI) events can help clinicians make effective decisions and establish appropriate therapy programs. Recently, many AI-based techniques have been proposed for predicting DDI associated events. However, most existing methods pay less attention to the potential correlations between DDI events and other multimodal data such as targets and enzymes. To address this problem, we propose a **M**ultimodal **D**eep **N**eural **N**etwork (MDNN) 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. Finally, a multimodal fusion neural layer is designed to explore the complementary among the drug multimodal representations. We conduct extensive experiments on real-world dataset. The results show that MDNN can accurately predict DDI events and outperform the state-of-the-art models.

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. 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 [Vilar *et al.*, 2014]. However, DDI can lead to different biological consequences and events. For example, drug *Itraconazole* and drug *Abemaciclib* interaction together cause an event that *the risk to increase due to the severity of the adverse effects*, as shown in Figure 1. Therefore, accurate prediction of DDI events becomes a clinically important task which could help clinicians make effective decisions and establish appropriate therapy programs.

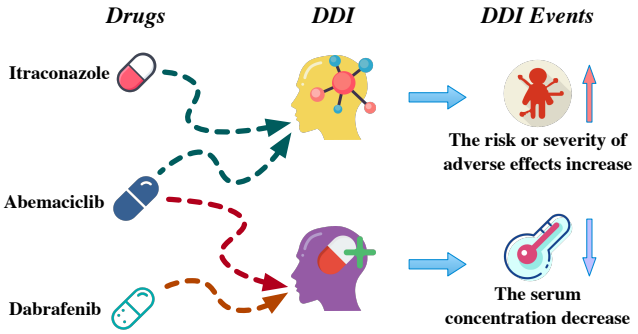

Figure 1: An example of DDI events. When drug *Abemaciclib* and drug *Dabrafenib* interaction together, an DDI event will be occurred and cause the decrease of body's serum concentration. However, it will raise the risk or severity of adverse effects when mixing drug *Abemaciclib* and drug *Itraconazole*.

The correct use of multiple drugs can minimize the medical risks while maximizing the synergy benefits of drugs.

There have been a number of AI-based models proposed for DDI events prediction, including analyzing chemical structure similarity using graph neural networks [Huang *et al.*, 2020], implementing multi-task learning on DDI type prediction [Jin *et al.*, 2017; Zitnik *et al.*, 2018; Ryu *et al.*, 2018], modeling semi-supervised learning to mine useful information for DDI prediction in both labeled and unlabeled drug data [Chu *et al.*, 2019], and exploiting knowledge graph summarization for multi-typed DDI pharmacological effect prediction [Yu *et al.*, 2021]. 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 [Ma *et al.*, 2018; Zhang *et al.*, 2015; Deng *et al.*, 2020]. 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 tackle the above limitations, this work aims 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 predic-

*Corresponding Author

tion. 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 networks that try to learn from structure information [Hamilton *et al.*, 2017; Wang *et al.*, 2019a; Cui *et al.*, 2020], 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 propose a new multimodal deep neural network with a two-pathway framework including the drug knowledge graph pathway and the heterogeneous feature 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 the state-of-the-art methods.

2 Related Work

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

Many efforts have been taken on calculating the similarities by integrating multiple data sources and predicting DDI based on the fused similarity. For example, the works [Vilar *et al.*, 2014; Abdelaziz *et al.*, 2017] integrate multiple drug features to calculate the similarities among drugs , and then predict DDI accurately based on the fused similarity. [Zhang *et al.*, 2015] proposes an integrative framework to fuse the similarities of drug features with proper weights and predict DDI. [Ma *et al.*, 2018] proposes to learn accurate and interpretable similarity measurement from multiple types of drug features for DDI prediction. In addition, [Deng *et al.*, 2020] proposes a framework DDIMDL that combines diverse drug features to build a model for predicting DDI events. However, they are limited in obtaining the rich features of drugs in structural information and semantic relations.

Recently, there has been growing interests in applying AI techniques for DDI prediction such as deep learning and graph neural networks. Different from drug similarity obtained from multiple sources, a deep learning framework named DeepDDI [Ryu *et al.*, 2018] is proposed to use molecular structures of drugs as inputs for predicting DDI types. The work [Jin *et al.*, 2017] proposes a new multitask dyadic prediction model to predict adverse drug-drug interactions. MLRDA [Chu *et al.*, 2019] develops a multi-task semi-supervised learning framework which effectively exploits information that is beneficial for DDI prediction in unlabeled drug data.

Inspired by the success of applying graph neural networks (GNN) in a wide variety of tasks [Jia *et al.*, 2020; Song *et al.*, 2020; Hao *et al.*, 2020], researchers also tried to utilize GNN to improve the performance of DDI events prediction. For example, Decagon [Zitnik *et al.*, 2018] applies a relational GNN for predicting side effects of drug pairs. [Yue *et al.*, 2020] integrates graph embedding methods for DDI prediction task. In addition, CASTER [Huang *et al.*, 2020] develops an end-to-end dictionary learning framework for predicting DDI with chemical structures of drugs. KGNN [Lin *et al.*, 2020] designs an effective framework for DDI prediction which can capture drug and its potential neighborhoods in the knowledge graph. Although these methods have achieved relatively good performance, they do not consider the drug multimodal data coherence and complementarity together. In addition, knowledge graph can provide a large amount of structured information among multiple entities and semantic relations associated with entities. Knowledge graphs are a powerful tool [Zhao *et al.*, 2020; Wang *et al.*, 2019b], 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 [Lin *et al.*, 2020]. However, most of these methods ignore the multimodal data. Moreover, only a few methods take different drug features as independent data and do not take cross-modality complementarity into consideration. 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.

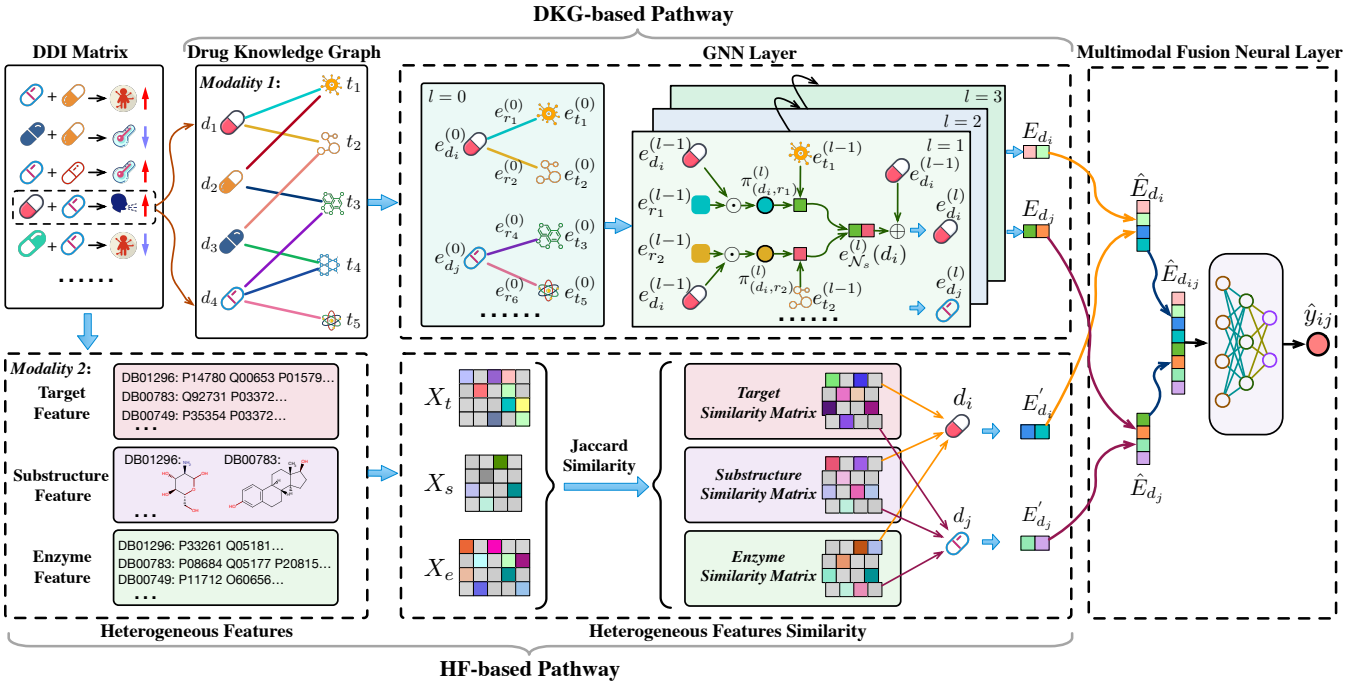

Figure 2: Illustration of the proposed MDNN, consisting of two core pathways: the DKG-based pathway and the HF-based pathway. (1) The DKG-based pathway utilizes the graph neural network to extract the topological structural information and semantic relations from the constructed drug knowledge graph (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.

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

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

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 features $\mathcal{X}_d$, 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.

4 Proposed Method

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

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

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:

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

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.

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:

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

where $e_{r_{in}}^{(l-1)}$ is the relation representation between drug d_i and tail entity t_n after $(l-1)^{th}$ GNN layer. $e_{d_i}^{(l-1)}$ 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:

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

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

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

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 $e_{d_i}^t \in E^t$, $e_{d_i}^s \in E^s$ and $e_{d_i}^e \in E^e$, 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:

$$E_{d_i}^{'} = e_{d_i}^t \oplus e_{d_i}^s \oplus e_{d_i}^e \tag{8}$$

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

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 $\hat{y}_{ij}$.

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.

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 [Ryu *et al.*, 2018]	0.8371	0.8899	0.9961	0.6848	0.7275	0.6611
DDIMDL [Deng *et al.*, 2020]	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**

Table 1: Performance of our model against competitive approaches. The best results are highlighted in boldface.

5 Experiments

In this section, we describe the experimental setups and the results of the performance evaluation on our proposed model in DDI events prediction.

5.1 Experimental Setup

Dataset. In order to demonstrate the effectiveness of our proposed model, we conduct extensive experiments on a real-world dataset including three parts: (1) **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. (2) **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. (3) **Heterogeneous Features:** Heterogeneous features released by DDIMDL [Deng *et al.*, 2020] include 1,162 target features, 583 substructure features and 202 enzyme features.

Baselines. In this study, we compare our model against the following baselines, including the traditional and the recent state-of-the-art methods. DDIMDL [Deng *et al.*, 2020] adopts a joint deep neural network framework to learn the representations of drug–drug pairs and predict DDI events. DeepDDI [Ryu *et al.*, 2018] develops a deep learning-based method that reduces the dimension of drug features based on a principal component analysis. 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.

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 $\mathcal{N}_s = 6$, $l = 1$, 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.

5.2 Results and Analysis

In this section, we report the performance of our model and all baselines in Table 1. From the result, we find that our model achieves the best performance in DDI events prediction on the real-world dataset. 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 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

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. 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. MDNN$_{hf}$ is the model variant where we only explore cross-modality embedding of drug–drug pairs using only the heterogeneous features. Moreover, as MDNN$_{hf}$ only considers multimodal attribute feature of drug–drug pairs, it performs worse than MDNN and MDNN$_{dkg}$ in all metrics. Figure 3 shows that the ablation results which verify the contribution of each pathway in our model, showing that combining the topological

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

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

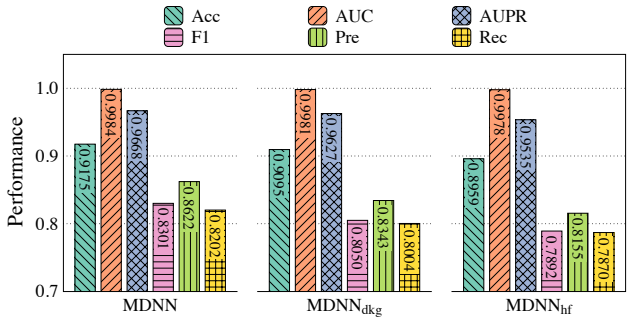

Figure 3: Results of the ablation experiments with the relative performances compared with complete MDNN in all the metrics. MDNN$_{dkg}$ and MDNN$_{hf}$ mean DKG-based and HF-based pathway embeddings, respectively.

representations in neighborhood with the semantic relations from the DKG and the heterogeneous features is beneficial to improving the DDI events prediction performance. It can be seen that from the results that MDNN outperforms both variants in all metrics.

5.4 Parameter Sensitivity Analysis

In this work, there are three essential parameters, which are the size of neighborhood sample $\mathcal{N}_s$, the number of GNN layers l and the dimension of embedding d in DKG. We fixed other parameters when studying the effect of each of them. The results are shown in Figure 4.

Effect of neighborhood size. We vary the size of sampled neighbor $\mathcal{N}_s$ to explore the efficacy of MDNN. Figure 4 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 GNN layers. We investigate the influence of the GNN layer l by varying its value from 1 to 3. 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 embedding dimension. In addition, we examine the influence of embedding dimension d by varying from its value from 32 to 512. Intuitively, the performance can be enhanced with a 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.

5.5 Multi-task Analysis

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 DDI between training drugs, and then make predictions for DDI events between training drugs and test drugs. For task B, it is different from task A that MDNN make predictions for DDI events between test drugs.

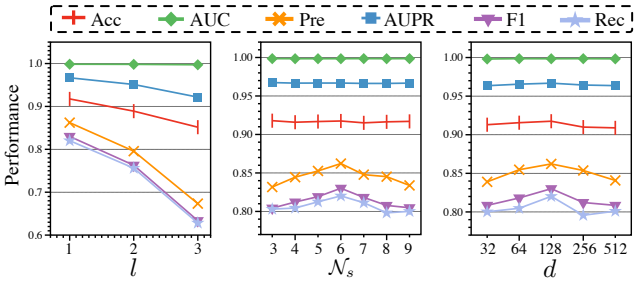

Figure 4: Results of MDNN with varying values of GNN layer l, neighborhood sample $\mathcal{N}_s$, and initialization dimension d in DKG.

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**

Table 2: Performance comparison of MDNN with other methods on two different tasks.

It can be learned from Table 2 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

In this paper, we propose a new MDNN model for drug-drug interaction events prediction. 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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61873288), Zhejiang Lab (2019KE0AB01, 111007-PI2001), Zhejiang Provincial Natural Science Foundation (LZ21F030001), Hunan Provincial Innovation Foundation for Postgraduate (CX20200108) and the Fundamental Research Funds for the Central Universities of Central South University (2021zzts0740).

References

[Abdelaziz *et al.*, 2017] Ibrahim Abdelaziz, Achille Fokoue, Oktie Hassanzadeh, Ping Zhang, and Mohammad Sadoghi. Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions. *Journal of Web Semantics*, 44:104–117, 2017.

[Chu *et al.*, 2019] Xu Chu, Yang Lin, Yasha Wang, Leye Wang, Jiangtao Wang, and Jingyue Gao. Mlrda: A multi-task semi-supervised learning framework for drug-drug interaction prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4518–4524, 2019.

[Cui *et al.*, 2020] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 492–502, 2020.

[Deng *et al.*, 2020] Yifan Deng, Xinran Xu, Yang Qiu, Jingbo Xia, Wen Zhang, and Shichao Liu. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*, 36:4316–4322, 2020.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, 2017.

[Hao *et al.*, 2020] Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang, Zheyuan Hu, Qi Liu, Enhong Chen, and Cheekong 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*, pages 731–752, 2020.

[Huang *et al.*, 2020] 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*, volume 34, pages 702–709, 2020.

[Jia *et al.*, 2020] Ziyu Jia, Youfang Lin, Jing Wang, Ronghao Zhou, Xiaojun Ning, Yuanlai He, and Yaoshuai Zhao. Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1324–1330, 2020.

[Jin *et al.*, 2017] Bo Jin, Haoyu Yang, Cao Xiao, Ping Zhang, Xiaopeng Wei, and Fei Wang. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 1367–1373, 2017.

[Lin *et al.*, 2020] Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, and Xiangxiang Zeng. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2739–2745, 2020.

[Ma *et al.*, 2018] Tengfei Ma, Cao Xiao, Jiayu Zhou, and Fei Wang. Drug similarity integration through attentive multi-view graph auto-encoders. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3477–3483, 2018.

[Ryu *et al.*, 2018] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. Deep learning improves prediction of drug-drug and drug-food interactions. *Proceedings of the National Academy of Sciences*, 115(18):E4304–E4311, 2018.

[Song *et al.*, 2020] Ying Song, Shuangjia Zheng, Zhangming Niu, Zhang-Hua Fu, Yutong Lu, and Yuedong Yang. Communicative representation learning on attributed molecular graphs. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2831–2838, 2020.

[Vilar *et al.*, 2014] Santiago Vilar, Eugenio Uriarte, Lourdes Santana, Tal Lorberbaum, George Hripcsak, Carol Friedman, and Nicholas P Tatonetti. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols*, 9(9):2147–2163, 2014.

[Wang *et al.*, 2019a] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. Knowledge graph convolutional networks for recommender systems. In *Proceedings of the World Wide Web Conference*, pages 3307–3313, 2019.

[Wang *et al.*, 2019b] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 950–958, 2019.

[Yu *et al.*, 2021] Yue Yu, Kexin Huang, Chao Zhang, Lucas M Glass, Jimeng Sun, and Cao Xiao. Sumgnn: Multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, pages 1–8, 2021.

[Yue *et al.*, 2020] Xiang Yue, Zhen Wang, Jingong Huang, Srinivasan Parthasarathy, Soheil Moosavinasab, Yungui Huang, Simon M Lin, Wen Zhang, Ping Zhang, and Huan Sun. Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics*, 36(4):1241–1251, 2020.

[Zhang *et al.*, 2015] Ping Zhang, Fei Wang, Jianying Hu, and Robert Sorrentino. Label propagation prediction of drug-drug interactions based on clinical side effects. *Scientific reports*, 5(1):1–10, 2015.

[Zhao *et al.*, 2020] Yang Zhao, Jiajun Zhang, Yu Zhou, and Chengqing Zong. Knowledge graphs enhanced neural machine translation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4039–4045, 2020.

[Zitnik *et al.*, 2018] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.