10-2024

H2D: Hierarchical Heterogeneous Graph Learning Framework for Drug-Drug Interaction Prediction

Ran Zhang
Chinese Academy of Sciences

Xuezhi Wang
Chinese Academy of Sciences, Beijing

Sheng Wang
University of Washington

Kunpeng Liu
Portland State University, kunpeng@pdx.edu

Yuanchun Zhou
Chinese Academy of Sciences

See next page for additional authors

Citation Details

Ran Zhang, Xuezhi Wang, Sheng Wang, Kunpeng Liu, Yuanchun Zhou, and Pengfei Wang. 2024. H2D: Hierarchical Heterogeneous Graph Learning Framework for Drug-Drug Interaction Prediction. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA. ACM, New York, NY, USA,

H2D: Hierarchical Heterogeneous Graph Learning Framework for Drug-Drug Interaction Prediction

Ran Zhang
Computer Network Information Center, Chinese Academy of Science
University of Chinese Academy of Sciences
Beijing, China
zhangran@cnic.cn

Xuezhi Wang
Computer Network Information Center, Chinese Academy of Science
Beijing, China
wxz@cnic.cn

Sheng Wang
University of Washington
Seattle, Washington, USA
swang@cs.washington.edu

Kunpeng Liu
Portland State University
Portland, Oregon, USA
kunpeng@pdx.edu

Yuanchun Zhou
Computer Network Information Center, Chinese Academy of Science
Beijing, China
zyc@cnic.cn

Pengfei Wang*
Computer Network Information Center, Chinese Academy of Science
Beijing, China
pfwang@cnic.cn

Abstract

Accurately predicting Drug-Drug Interactions (DDIs) is critical to designing effective drug combination therapies. Recently, Artificial Intelligence (AI)-powered DDI prediction approaches have emerged as a new paradigm. However, most existing methods oversimplify the complex hierarchical structure within molecules and overlook the multi-source heterogeneous information external to molecules, limiting their modeling and predictive capabilities. To address this, we propose a **H**ierarchical **H**eterogeneous graph learning framework for **D**DI prediction, namely **H2D**. H2D employs an internal-to-external, local-to-global hierarchical perspective, exploiting intra-molecular multi-granularity structures and inter-molecular biomedical interactions to mutually enhance across hierarchical levels. Extensive experimental results demonstrate H2D's effectiveness on three real-world DDI prediction tasks (binary-class, multi-class, and multi-label). In sum, H2D achieves state-of-the-art performance in DDI prediction by leveraging the multi-scale graph structures, opening up new avenues in AI-powered DDI prediction.

CCS Concepts

• **Applied computing** → Molecular structural biology; • **Computing methodologies** → Neural networks.

Keywords

Drug-drug interaction, Hierarchical graph learning, Heterogeneous graph neural network

*Corresponding author.

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679936

ACM Reference Format:
Ran Zhang, Xuezhi Wang, Sheng Wang, Kunpeng Liu, Yuanchun Zhou, and Pengfei Wang. 2024. H2D: Hierarchical Heterogeneous Graph Learning Framework for Drug-Drug Interaction Prediction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3627673.3679936

1 Introduction

Drug-Drug Interactions (DDIs) represent a significant concern in clinical pharmacology and healthcare, as they profoundly influence the safety and efficacy of medication regimens [5, 13]. As patients are increasingly prescribed multiple medications to manage complex medical conditions, the risk of drug interactions grows [10]. Therefore, identifying potential DDIs is crucial for avoiding adverse drug reactions and optimizing clinical decision-making. Traditionally, DDI prediction relies on empirical studies and clinical trials, which are time-consuming, expensive, and often limited due to the vast combinatorial space of potential DDIs [21]. With advancements of deep learning and Artificial Intelligence (AI) technology [6, 19, 28], AI-powered methods have emerged as promising alternatives for DDI prediction [1, 2, 25, 29].

AI-powered DDI prediction methods can be categorized into intra-level, inter-level, and multi-level approaches. Specifically, intra-level methods focus on utilizing molecular characteristics such as chemical sequences and structures [9, 12, 14, 18, 23, 27]. Inter-level methods integrate interactions between molecules as well as relationships with other entities, such as genes, pathways, and diseases [11, 15, 26, 30]. Multi-level methods incorporate both internal properties and external associations related to DDIs [3, 8].

Although the above works have achieved encouraging results, they still face the following challenges: ***(1) How to leverage multi-scale molecular structures for enhancing DDI prediction?*** Atoms are the basic particles of molecules, while motifs represent frequently recurring subgraph patterns that convey semantic meanings [24]. Most DDI prediction methods often concentrate on either fine-grained atom-level details or coarse-grained motif-level

information, falling short in providing a comprehensive understanding of drug structures. ***(2) How can the external heterogeneous network benefit from the internal molecular structure?*** Some methods consider both molecular graphs and biomedical networks but overlook dependencies among hierarchical levels. They simply concatenate the features learned independently from each level for prediction, limiting the synergistic potential of combining local molecular insights with global biomedical information. ***(3) How to make the model more adaptable to real-world scenarios?*** Most models only determine the presence of medication interactions without delving into the specific adverse consequences resulting from DDIs, which is crucial for clinical decision-making.

Along this line, we propose a **H**ierarchical **H**eterogeneous graph learning framework for **D**DI prediction (H2D). H2D provides a comprehensive understanding of molecules from an *atom-motif-interaction* perspective for effective binary-class, multi-class, and multi-label DDI prediction tasks. To summarize, H2D exhibits the following highlights and advantages:

- **Hierarchical perception from atom to motif to interaction.** H2D captures multi-scale structures of molecular graphs and multi-hop dependencies within biomedical networks, gaining multi-granularity molecular knowledge.
- **Interaction between multi-granularity information**. A set of micro-level graphs are interconnected by edges in a macro-level graph. H2D achieves sequential dependency and mutual enhancement across hierarchical levels.
- **Superior performance across diverse tasks**. H2D exhibits superior performance across three real-world tasks and demonstrates high adaptability in predicting DDIs involving multiple types of interactions in clinical applications.

2 Methodology

2.1 Problem Definition

A heterogeneous graph, denoted as $G = (V, E)$, is a graph with multiple node types $V = \{V_1, V_2, ..., V_m\}$ and multiple edge types $E = \{E_1, E_2, ..., E_n\}$. Each edge can be described as (v_i, e_k, v_j), where v_i and v_j are nodes of type V_i and V_j respectively, and e_k is an edge of type E_k. Given a ***biomedical interaction graph*** $\mathcal{G}_{bn} = (\mathcal{V}_{bn}, \mathcal{E}_{bn})$, where $\mathcal{V}_{bn}$ is the entity set, $\mathcal{V}_d \subset \mathcal{V}_{bn}$ is the drug set, and $\mathcal{E}_{bn}$ is the biomedical interaction set. For a node of drug set in the biomedical interaction graph, we define a molecule as a ***motif graph*** $\mathcal{G}_m = (\mathcal{V}_m, \mathcal{E}_m)$, where $\mathcal{V}_m$ represents the set of motifs within the molecule and $\mathcal{E}_m$ is the set of inter-motif bonds. Similarly, a node in the motif graph can be defined as an ***atom graph*** $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$, where $\mathcal{V}_a$ denotes the set of atoms within the motif and $\mathcal{E}_a$ represents the set of inter-atom bonds.

In our study, **DDI prediction** is defined as binary-class, multi-class, and multi-label link prediction tasks. We aim to use the hierarchical graph sets as inputs to generate the predicted results, indicating whether (binary-class and multi-label) or what type of interactions (multi-class) occur between drug pairs.

2.2 Framework Overview

As shown in Figure 1, H2D utilizes an *atom-motif-interaction* graph learning framework, effectively integrating internal molecular features with external relational information. Specifically, within the atom graph, we introduce a heterogeneous graph encoder to capture the structural arrangements of atoms within motifs. In the motif graph, we explore the relationships between motifs connected by diverse chemical bonds to understand inter-motif structures. For the interaction graph, we employ a heterogeneous graph attention network to aggregate dependencies and correlations among biomedical entities. Furthermore, we facilitate mutual enhancement between the intra-molecular and inter-molecular levels. Finally, the multi-level molecular embeddings are concatenated and fed into DDI predictors to perform binary-class, multi-class, and multi-label DDI predictions across three real-world scenarios.

2.3 Atom Graph

The internal structures of molecules are intricate, consisting of recurring and distinct substructures known as motifs. We employ the Simplified Molecular Input Line Entry System (SMILES) [20] as molecular descriptors and use the Breaking Recurrent Internal Chemical Structures (BRICS) [4] algorithm to decompose SMILES into chemically meaningful motifs by selectively breaking bonds according to predefined rules. To capture the characteristics of these motifs, we first focus on the corresponding atom graphs.

In the atom graph, we initialize node features for each atom based on its nuclear charge number and edge features for bonds based on chemical bond types, including single, double, triple, and aromatic bonds. By applying linear transformations, we project atom features and bond features into a hidden vector space $\mathbf{H}$. We obtain the initial state $\mathbf{h}_i$ for atom i, and $\mathbf{h}_{(i,j)}$ for bond (i, j). Following this, we introduce the Heterogeneous Graph Attention Network (HGAT) with residual connection to capture atom arrangements and encode complex structural details. The atom embedding at the $(l+1)$-th layer can be updated as follows:

$$\alpha_{ij} = \sigma\left(\mathbf{A}^{\mathbf{T}}\left[\mathbf{W}_{\varphi(i)}\mathbf{h}_i^{(l)} \,\middle\|\, \mathbf{W}_{\varphi(j)}\mathbf{h}_j^{(l)} \,\middle\|\, \mathbf{W}_{\psi(i,j)}\mathbf{h}_{(i,j)}\right]\right) \quad (1)$$

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{h}_i^{(l)} + \sum_{j\in\mathcal{N}_i} \alpha_{ij}\mathbf{W}_{\varphi(j)}\mathbf{h}_j^{(l)}\right) \quad (2)$$

where $\varphi(i)$ and $\psi(i, j)$ represent the types of atom i and bond (i, j), $\mathbf{A}^{\mathbf{T}}$, $\mathbf{W}_{\varphi(i)}$, and $\mathbf{W}_{\psi(i,j)}$ denote the learnable matrices, $\|$ denotes a concatenation operation, $\mathcal{N}_i$ is the neighborhood set of atom i, and σ is an activation function.

After L_1 iterations of HGAT block, we aggregate the embeddings of all atoms within the motif using a pooling function to obtain the motif embedding $\mathbf{m} = \text{Pooling}(\{\mathbf{h}_i^{(L_1)}\})$.

2.4 Motif Graph

A set of motifs (atom graphs) are interconnected within a molecule (motif graph) via inter-motif bonds. We utilize embeddings $\{\mathbf{m}_i\}$ derived from encoding atom graphs to initialize the node features of the motif graph, thereby incorporating detailed atom arrangements into motif interactions. We preserve the chemical bonds broken by BRICS, reconstruct the inter-motif relationships to establish motif graphs, and subsequently introduce the Graph Isomorphism Network (GIN) to capture the relative spatial arrangements among motifs, encoding complex structural information. At the $(l+1)$-th layer of GIN block, the motif embedding $\mathbf{m}_v^{(l+1)}$ is updated by aggregating information from its neighborhood $\mathcal{N}_v$ as follows:

$$\mathbf{m}_v^{(l+1)} = \text{MLP}^{(l)}\left((1+\epsilon^{(l)})\cdot\mathbf{m}_v^{(l)} + \sum_{u\in\mathcal{N}_v}\mathbf{m}_u^{(l)}\right) \quad (3)$$

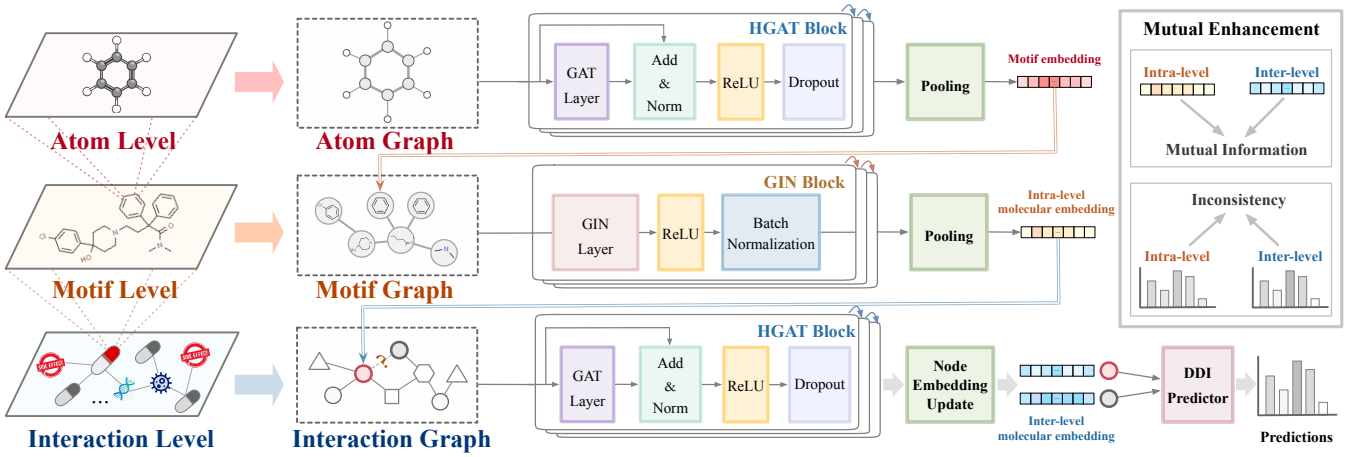

Figure 1: Overall framework of H2D. H2D is a hierarchical heterogeneous graph learning framework for DDI prediction, with three hierarchical and interdependent levels: atom level, motif level, and interaction level.

where $\mathbf{m}_v^{(0)} = \mathbf{m}_v$, $\mathcal{N}_v$ denotes the neighboring motif set of motif v, $\epsilon^{(l)}$ is a learnable scalar parameter used for message passing normalization, and $\text{MLP}^{(l)}$ is a Multi-Layer Perceptron (MLP).

We apply L_2 layers of GIN block, followed by a pooling function to obtain the intra-level molecular embedding $\mathbf{z} = \text{Pooling}\left(\{\mathbf{m}_v^{(L_2)}\}\right)$.

2.5 Interaction Graph

In the biomedical network, drug molecules are not isolated entities; they exhibit intricate associations with other molecules and entities such as diseases and pathways. We initialize the molecular feature $\mathbf{z}_i^{(0)}$ using intra-level molecular embedding $\mathbf{z}_i$, thereby integrating the molecular structural features into the biomedical network. Node embeddings for other entity types are randomly initialized. In addition, edge embeddings are determined by the types of interactions. Given the diverse significance of different interaction types in comprehending DDI mechanisms, we employ HGAT to aggregate the multi-hop dependencies of molecules, capturing potential semantics in the biomedical network as follows:

$$\beta_{ij} = \sigma\left(\mathbf{B}^{\mathbf{T}}\left[\mathbf{W}_{\gamma(i)}\mathbf{z}_i^{(l)} \,\middle\|\, \mathbf{W}_{\gamma(j)}\mathbf{z}_j^{(l)} \,\middle\|\, \mathbf{W}_{\delta(i,j)}\mathbf{z}_{\delta(i,j)}\right]\right) \tag{4}$$

$$\mathbf{z}_i^{(l+1)} = \sigma\left(\mathbf{z}_i^{(l)} + \sum_{j\in\mathcal{N}_i} \beta_{ij}\mathbf{W}_{\gamma(j)}\mathbf{z}_j^{(l)}\right) \tag{5}$$

where $\gamma(i)$ and $\delta(i,j)$ denote the types of entity i and interaction (i,j), $\mathcal{N}_i$ is the set of associated entities with entity i, $\mathbf{B}^{\mathbf{T}}$, $\mathbf{W}_{\gamma(i)}$ and $\mathbf{W}_{\delta(i,j)}$ are the learnable matrices.

Following information propagation over L_3 hops, we derive the inter-level molecular embedding, denoted as $\mathbf{g}_i = \mathbf{z}_i^{(L_3)}$, which incorporates the multi-source biomedical knowledge.

2.6 DDI Prediction

2.6.1 DDI Predictor. From the H2D framework, we obtain the intra- and inter-level embeddings of molecule i: $\mathbf{z}_i$ and $\mathbf{g}_i$. We then concatenate the two representations to obtain the final molecular representation $\mathbf{f}_i = [\mathbf{z}_i \,\|\, \mathbf{g}_i]$. We adopt an MLP to compress a pair of drug embeddings into a link value p_{ij}, signifying the interaction probability between drug i and drug j. DDI predictors generate three prediction scores: p_{ij}^z, p_{ij}^g, and p_{ij}^f, based on intra-level, inter-level, and multi-level drug representations, respectively. During testing, we use p_{ij}^f for the final DDI prediction.

2.6.2 Loss Function. The loss consists of two components: supervised loss and unsupervised loss. Specifically, supervised loss facilitates accurate DDI prediction, and unsupervised loss promotes mutual learning between intra- and inter-levels.

Supervised loss $\mathcal{L}_S$ measures the Cross Entropy (CE) between the predicted probability distribution and the true label.

$$\mathcal{L}_S = \sum_{(i,j)\in\Omega} \text{CE}\left(p_{ij}^z \,\|\, y_{ij}\right) + \text{CE}\left(p_{ij}^g \,\|\, y_{ij}\right) + \text{CE}\left(p_{ij}^f \,\|\, y_{ij}\right) \tag{6}$$

where Ω is the set of observed drug pairs in the training set, and y_{ij} denotes the true label of drug pair (i,j).

Unsupervised loss $\mathcal{L}_U$ quantifies the shared information between intra-level and inter-level molecular embeddings and calculates the Kullback-Leibler (KL) divergence between two DDI predictions derived from intra- and inter-levels.

$$\mathcal{L}_U = \lambda_1 \sum_{i\in\mathcal{V}_d} \text{MI}\left(\mathbf{g}_i \,\|\, \mathbf{z}_i\right) + \lambda_2 \sum_{(i,j)\in\Omega} \text{KL}\left(p_{ij}^g \,\|\, p_{ij}^z\right) \tag{7}$$

where λ_1 and λ_2 are trade-off hyper-parameters, $\mathcal{V}_d$ is the drug set, and MI represents the computation operator of mutual information.

Total loss $\mathcal{L}$ can be defined as the sum of supervised loss $\mathcal{L}_S$ and unsupervised loss $\mathcal{L}_U$, formulated as $\mathcal{L} = \mathcal{L}_S + \mathcal{L}_U$.

3 Experiment

In this section, we conduct extensive experiments and try to answer the following two research questions: **RQ1:** Can H2D improve DDI prediction performance in real-world tasks? **RQ2:** How do the different levels of H2D contribute to its overall performance?

Table 1: Overall DDI prediction performance on DrugBank dataset. The best results are highlighted in bold, and the runner-up results are highlighted in underline. (Higher values indicate better performance.)

Type	Model	Binary-class					Multi-class				
		Accuracy	Precision	Recall	F1	AUROC	Accuracy	Precision	Recall	F1	Kappa
Intra-level	DeepDDI	91.12	89.89	92.91	91.37	97.27	85.56	90.54	81.11	72.77	82.22
	Molormer	97.05	96.32	97.91	97.11	99.67	96.77	94.87	92.45	93.91	96.17
	MDF-SA-DDI	93.59	92.64	94.23	93.44	98.48	93.13	95.97	88.17	91.29	92.94
Inter-level	GAT	87.54	87.87	87.10	87.48	94.66	77.06	58.75	76.82	61.09	72.93
	KGNN	92.75	92.99	92.98	92.97	97.31	92.58	79.94	73.77	75.92	91.17
Multi-level	MUFFIN	96.69	96.34	97.08	96.71	99.47	96.96	94.53	92.38	93.08	96.54
	HetDDI	98.82	98.52	99.12	98.82	99.87	98.13	96.04	96.27	96.17	97.78
	H2D	**98.85**	**98.58**	**99.13**	**98.86**	**99.88**	**98.23**	**96.75**	**96.76**	**96.60**	**97.89**

Table 2: DDI prediction performance on Twosides dataset.

Model	Multi-label				
	Accuracy	Precision	Recall	F1	AUROC
DeepDDI	87.78	86.63	89.30	87.94	94.61
KGNN	92.09	93.30	90.71	91.99	97.55
MUFFIN	95.18	93.42	97.20	95.28	98.88
Molormer	94.81	92.40	97.60	94.93	98.74
HetDDI	96.66	96.08	97.29	96.68	99.34
H2D	**97.32**	**96.47**	**98.25**	**97.35**	**99.46**

3.1 Experimental Setups

3.1.1 Datasets. We evaluate H2D on two real-world datasets, DrugBank[1] [22] and Twosides[2] [16]. We use DRKG[3] [7] as the external biomedical network, which consists of 97,238 entities with 13 node types and 5,874,261 interactions with 107 edge types. Following preprocessing, DrugBank contains 1706 drugs and 191,427 DDIs with 86 interaction types. Twosides contains 1,345 drugs and 1,979,575 DDIs with 200 interaction types.

3.1.2 Evaluation Metrics. We select six widely used metrics: Accuracy, Precision, Recall, F1-score, Area Under the Receiver Operating Characteristic curve (AUROC), and Kappa. We report the mean results of five-fold cross-validation experiments.

3.1.3 Baselines. We compare H2D with three representative categories of baselines as follows: (1) *intra-level models*: DeepDDI [14], Molormer [27], and MDF-SA-DDI [9]; (2) *inter-level models*: GAT [17] and KGNN [11]; (3) *multi-level models*: MUFFIN [3] and HetDDI [8].

3.1.4 Experimental Settings. All experiments are conducted on the Linux server with EPYC 7742 CPU and TESLA A100 GPU. We set $L_1 = L_2 = L_3 = 3$, the learning rate as 0.001.

3.2 Performance Analysis

Overall Performance (RQ1). Table 1 and Table 2 list the experimental results of all methods across binary-class, multi-class, and multi-label DDI prediction tasks. In general, multi-level methods achieve better performance compared to single-level methods, illustrating the importance of integrating both intra-level and inter-level molecular information. Furthermore, the experimental results indicate that H2D outperforms all baselines in three real-world tasks, demonstrating the effectiveness of the hierarchical graph learning framework and heterogeneous graph encoders in H2D. In particular, H2D exhibits notable performance improvements in multi-class and multi-label DDI prediction tasks, showcasing its strong adaptability in clinical applications.

Ablation Study (RQ2). H2D is a hierarchical graph learning framework that integrates atom, motif, and interaction levels. To investigate the contribution of each level to model performance, we conduct an ablation study on hierarchical graphs in the multi-class task. As depicted in Figure 2, incorporating hierarchical graphs greatly enhances DDI prediction performance. The atom level and motif level provide complementary knowledge, equipping H2D with sensitivity to multi-granularity intra-molecular structures. Additionally, in the multi-class task, biomedical interactions play a more significant role than molecular structural information, highlighting the importance of exploring high-order dependencies and long-range correlations in biomedical networks.

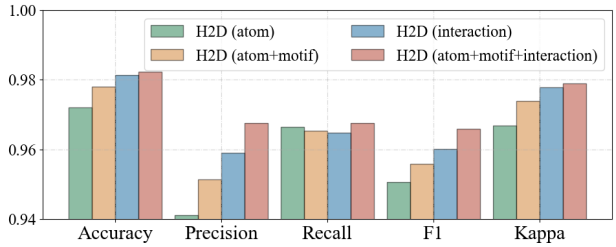

Figure 2: Ablation experimental results on different levels.

4 Conclusion

In this paper, we propose H2D, an *atom-motif-interaction* graph learning framework for DDI prediction. H2D integrates multi-granularity structures within molecules, explores multi-type interactions beyond molecules, and achieves mutual enhancement across hierarchical levels. Extensive experimental results validate the superior performance of H2D across real-world tasks.

Acknowledgments

This work is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences XDB38030300.

[1]https://bitbucket.org/kaistsystemsbiology/deepddi/src/master/data/
[2]https://tatonettilab.org/offsides/
[3]https://github.com/gnn4dr/DRKG/

References

[1] Heba Askr, Enas Elgeldawi, Heba Aboul Ella, Yaseen AMM Elshaier, Mamdouh M Gomaa, and Aboul Ella Hassanien. 2023. Deep learning in drug discovery: an integrative review and future challenges. *Artificial Intelligence Review* 56, 7 (2023), 5975–6037.

[2] Siqi Chen, Ivan Semenov, Fengyun Zhang, Yang Yang, Jie Geng, Xuequan Feng, Qinghua Meng, and Kaiyou Lei. 2024. An effective framework for predicting drug–drug interactions based on molecular substructures and knowledge graph neural network. *Computers in Biology and Medicine* 169 (2024), 107900.

[3] Yujie Chen, Tengfei Ma, Xixi Yang, Jianmin Wang, Bosheng Song, and Xiangxiang Zeng. 2021. MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* 37, 17 (2021), 2651–2658.

[4] Jorg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. 2008. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem* 3, 10 (2008), 1503.

[5] Kathleen M Giacomini, Ronald M Krauss, Dan M Roden, Michel Eichelbaum, Michael R Hayden, and Yusuke Nakamura. 2007. When good drugs go bad. *Nature* 446, 7139 (2007), 975–977.

[6] Bruna Gomes and Euan A Ashley. 2023. Artificial intelligence in molecular medicine. *New England Journal of Medicine* 388, 26 (2023), 2456–2465.

[7] Vassilis N Ioannidis, Da Zheng, and George Karypis. 2020. Few-shot link prediction via graph neural networks for covid-19 drug-repurposing. *arXiv preprint arXiv:2007.10261* (2020).

[8] Zhe Li, Xinyi Tu, Yuping Chen, and Wenbin Lin. 2023. HetDDI: a pre-trained heterogeneous graph neural network model for drug–drug interaction prediction. *Briefings in Bioinformatics* 24, 6 (2023), bbad385.

[9] Shenggeng Lin, Yanjing Wang, Lingfeng Zhang, Yanyi Chu, Yatong Liu, Yitian Fang, Mingming Jiang, Qiankun Wang, Bowen Zhao, Yi Xiong, et al. 2022. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics* 23, 1 (2022), bbab421.

[10] Xuan Lin, Lichang Dai, Yafang Zhou, Zu-Guo Yu, Wen Zhang, Jian-Yu Shi, Dong-Sheng Cao, Li Zeng, Haowen Chen, Bosheng Song, et al. 2023. Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction. *Briefings in Bioinformatics* 24, 4 (2023), bbad235.

[11] Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, and Xiangxiang Zeng. 2020. KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction.. In *IJCAI*, Vol. 380. 2739–2745.

[12] Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. 2021. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics* 22, 6 (2021), bbab133.

[13] Thomayant Prueksaritanont, Xiaoyan Chu, Christopher Gibson, Donghui Cui, Ka Lai Yee, Jeanine Ballard, Tamara Cabalu, and Jerome Hochman. 2013. Drug–drug interaction studies: regulatory guidance and an industry perspective. *The AAPS journal* 15 (2013), 629–645.

[14] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences* 115, 18 (2018), E4304–E4311.

[15] Xiaorui Su, Zhuhong You, Deshuang Huang, Lei Wang, Leon Wong, Boya Ji, and Bowei Zhao. 2022. Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction. *IEEE Transactions on Knowledge and Data Engineering* 35, 6 (2022), 5640–5651.

[16] Nicholas P Tatonetti, Patrick P Ye, Roxana Daneshjou, and Russ B Altman. 2012. Data-driven prediction of drug effects and interactions. *Science translational medicine* 4, 125 (2012), 125ra31–125ra31.

[17] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat* 1050, 20 (2017), 10–48550.

[18] Jinxian Wang, Xuejun Liu, Siyuan Shen, Lei Deng, and Hui Liu. 2022. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics* 23, 1 (2022), bbab390.

[19] Zaitian Wang, Pengfei Wang, Kunpeng Liu, Pengyang Wang, Yanjie Fu, Chang-Tien Lu, Charu C Aggarwal, Jian Pei, and Yuanchun Zhou. 2024. A Comprehensive Survey on Data Augmentation. *arXiv preprint arXiv:2405.09591* (2024).

[20] David Weininger. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* 28, 1 (1988), 31–36.

[21] Steven Whitebread, Jacques Hamon, Dejan Bojanic, and Laszlo Urban. 2005. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug discovery today* 10, 21 (2005), 1421–1433.

[22] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46, D1 (2018), D1074–D1082.

[23] Nuo Xu, Pinghui Wang, Long Chen, Jing Tao, and Junzhou Zhao. 2019. Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. *arXiv preprint arXiv:1905.09558* (2019).

[24] Xuan Zang, Xianbing Zhao, and Buzhou Tang. 2023. Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry* 6, 1 (2023), 34.

[25] Ran Zhang, Xuezhi Wang, Pengfei Wang, Zhen Meng, Wenjuan Cui, and Yuanchun Zhou. 2023. HTCL-DDI: a hierarchical triple-view contrastive learning framework for drug–drug interaction prediction. *Briefings in Bioinformatics* 24, 6 (2023), bbad324.

[26] Ran Zhang, Zhanjie Wang, Xuezhi Wang, Zhen Meng, and Wenjuan Cui. 2023. MHTAN-DTI: Metapath-based hierarchical transformer and attention network for drug–target interaction prediction. *Briefings in Bioinformatics* 24, 2 (2023), bbad079.

[27] Xudong Zhang, Gan Wang, Xiangyu Meng, Shuang Wang, Ying Zhang, Alfonso Rodriguez-Paton, Jianmin Wang, and Xun Wang. 2022. Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug–drug interactions prediction. *Briefings in Bioinformatics* 23, 5 (2022), bbac296.

[28] Xinhao Zhang, Zaitian Wang, Lu Jiang, Wanfu Gao, Pengfei Wang, and Kunpeng Liu. 2024. TFWT: Tabular Feature Weighting with Transformer. *arXiv preprint arXiv:2405.08403* (2024).

[29] Yuanyuan Zhang, Zengqian Deng, Xiaoyu Xu, Yinfei Feng, and Shang Junliang. 2023. Application of Artificial Intelligence in Drug–Drug Interactions Prediction: A Review. *Journal of Chemical Information and Modeling* (2023).

[30] Chengshuai Zhao, Shuai Liu, Feng Huang, Shichao Liu, and Wen Zhang. 2021. CSGNN: Contrastive Self-Supervised Graph Neural Network for Molecular Interaction Prediction.. In *IJCAI*. 3756–3763.