OXFORD

Briefings in Bioinformatics, 2023, **24(4)**, 1–20
https://doi.org/10.1093/bib/bbad235
Advance access publication date 3 July 2023
Review

Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction

Xuan Lin, Lichang Dai, Yafang Zhou, Zu-Guo Yu, Wen Zhang, Jian-Yu Shi, Dong-Sheng Cao, Li Zeng, Haowen Chen, Bosheng Song, Philip S. Yu and Xiangxiang Zeng

Corresponding author. Haowen Chen, College of Computer Science and Electronic Engineering, Hunan University, 410013 Changsha, P. R. China. Tel/Fax: +86 18932430631; E-mail: hwchen@hnu.edu.cn; Bosheng Song, College of Computer Science and Electronic Engineering, Hunan University, 410013 Changsha, P. R. China. E-mail:boshengsong@hnu.edu.cn

Abstract

Recent advances and achievements of artificial intelligence (AI) as well as deep and graph learning models have established their usefulness in biomedical applications, especially in drug–drug interactions (DDIs). DDIs refer to a change in the effect of one drug to the presence of another drug in the human body, which plays an essential role in drug discovery and clinical research. DDIs prediction through traditional clinical trials and experiments is an expensive and time-consuming process. To correctly apply the advanced AI and deep learning, the developer and user meet various challenges such as the availability and encoding of data resources, and the design of computational methods. This review summarizes chemical structure based, network based, natural language processing based and hybrid methods, providing an updated and accessible guide to the broad researchers and development community with different domain knowledge. We introduce widely used molecular representation and describe the theoretical frameworks of graph neural network models for representing molecular structures. We present the advantages and disadvantages of deep and graph learning methods by performing comparative experiments. We discuss the potential technical challenges and highlight future directions of deep and graph learning models for accelerating DDIs prediction.

Keywords: deep learning, graph learning, drug–drug interactions prediction

INTRODUCTION

Polypharmacy is progressively becoming the prevalent therapy by a patient for one or more conditions, especially for older patients with many chronic health conditions, and this trend continues to grow because of aging populations. For example, 67% of elderly Americans were taking five or more medications [1]. This can be baffling because potential drug interactions can alter the intended responses when patients taking multiple drugs simultaneously, which results in unexpected side effects or decreases clinical efficacy [2]. These unintended interactions are widely referred to as drug–drug interactions (DDIs). As a common problem during polypharmacy, DDIs are associated with about 30% of all reported adverse drug effects that becomes one of the most leading causes of trial failures in drug discovery and clinical research [3, 4]. Take Ondansetron (Zofran) and dofetilide (Tikosyn) as an example. The former is a medication used to prevent nausea and vomiting, and the latter is used for heart rhythm. When they are used together, the amount of time between heartbeats can get too long. This can lead to dizziness, fainting and even death in severe cases. As a result, predicting potential DDIs in advance is crucial for drug development and pharmacovigilance.

Xuan Lin is currently a lecturer in the college of computer science, Xiangtan University , Xiangtan, China. His main research interests include machine learning, graph neural networks and bioinformatics.
Lichang Dai is an undergraduate in the college of computer science, Xiangtan University. His research interests are machine learning and bioinformatics.
Yafang Zhou is an undergraduate in the college of computer science, Xiangtan University. Her research interests are machine learning and data mining.
Zu-Guo Yu is a professor in Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University. His research is focused on fractals, bioinformatics and complex networks and geomagnetic data analysis.
Wen Zhang is a professor in College of Informatics, Huazhong Agricultural University, China. His research interests include machine learning and bioinformatics.
Jian-Yu Shi is a Professor in the Northwestern Polytechnical University, Xian, China. His research interests include bioinformatics, cheminformatics and artificial intelligence.
Dong-Sheng Cao is currently a professor in the Xiangya School of Pharmaceutical Sciences, Central South University, China. His research interests can be found at the website of his group: http://www.scbdd.com.
Li Zeng is currently the head of AIDD department of Yuyao Biotech, Shanghai, China. He has published several research papers on Biochemical Pharmacology, The FASEB Journal, Scientific Reports, etc.
Haowen Chen is an associate professor with Hunan University. His research interests include bioinformatics and artificial intelligence. He has published several research papers in these fields including Npj Systems Biology, IEEE/ACM TCBB, Methods, COMPUTERS & SECURITY etc.
Bosheng Song is an associate professor with the College of Information Science and Engineering, Hunan University, Changsha, China. His current research interests include membrane computing and bioinformatics.
Philip S. Yu (Fellow, IEEE) is a distinguished professor in computer science with the University of Illinois at Chicago and also holds the Wexler chair in information technology. His research interest is on Big Data, including data mining, data stream, database and privacy.
Xiangxiang Zeng (Senior Member, IEEE) is an Yuelu distinguished professor with the College of Information Science and Engineering, Hunan University, Changsha, China. His main research interests include computational intelligence, graph neural networks and bioinformatics.
Received: March 27, 2023. **Revised:** May 30, 2023. **Accepted:** June 5, 2023
© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

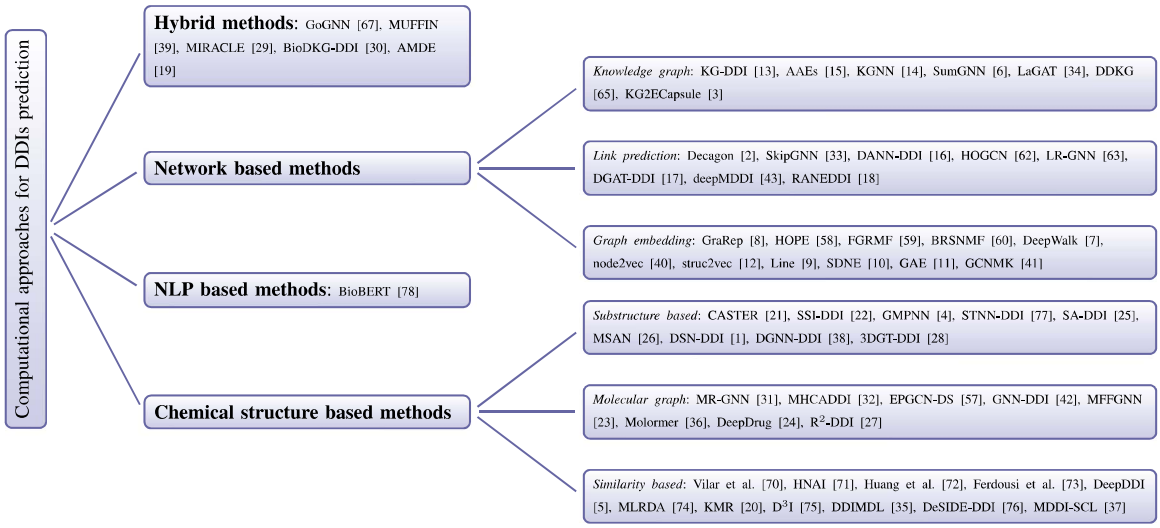

Figure 1. The taxonomy of computational methods with representative examples. The computational approaches can be divided into chemical structure based, network based, NLP based and hybrid method, respectively. *Chemical structure based methods* learn the structural and chemical features of molecules by using the similarity functions. These methods can be further classified into similarity based, molecular graph based and substructure based methods. *Network based methods* that mainly rely on the topological characteristics of drugs in related networks integrate biological information from different data sources, which can be further classified into graph embedding, link prediction and knowledge graph based methods. *Natural language processing (NLP) based methods* adopt the unsupervised and pretraining strategies in NLP to learn contextual information from large unlabeled molecular datasets. *Hybrid methods* combine with multiple types of features in an efficient pattern.

Identifying the existence of DDIs is the first step to avoid the potential adverse effects. Generally, DDIs can be broadly categorized into pharmaceutical, pharmacokinetic (PK) or pharmacodynamic (PD). Importantly, DDIs that primarily cause a change in PK will consequently lead to a secondary alteration in its PD. Thus, classifying DDIs' types is the further study that typically has been performed through extensive experimental testing in pharmaceutical research [5], which can help the scientific communities and manufacturers further decrease toxicity and increase effectiveness for these interactions [6]. However, more direct costs are incurred during long-period of clinical trials. Besides, DDIs have important relation with combination therapies that are hard to explore the space of combinations via high-throughput screening due to the exceedingly large number of unique chemical combinations. Collectively, these considerations highlight the benefits of discovering novel computational methods for predicting DDIs.

In the past few years, people have seen a surge in DDIs research due to the unprecedented success of deep and graph learning. Various databases, approaches and models have been proposed in the recent literature, urgently calling for a comprehensive survey to focus the efforts in this flourishing new direction. Few review articles cover machine learning algorithms and recently developed deep and graph learning models. Some surveys of databases and other resources supporting drug discovery and DDIs extraction have also been conducted recently [7–12]. In this survey, we first summarize the commonly adopted databases and molecular representations related to DDIs. Then we present an overview of computational methods for DDIs prediction and focus on reviewing deep learning and graph neural network (GNN) based methods. Moreover, we introduce several commonly-used GNN models. Furthermore, we select several representations baselines for comparative experiments on two benchmark datasets, and the detailed experimental results are analyzed. Finally, we make a conclusion to discuss the potential future trends as well as promising research directions that could be used to further improving DDIs prediction. To summarize, the main contributions of this work are as follows:

(i) *Structured taxonomy*. As shown in Figure 1, we contribute a structured taxonomy to provide a broad overview of computational methods, which categorizes existing works from four perspectives: chemical structure based, network based, NLP based and hybrid methods.
(ii) *Current progress*. We systematically delineate the current research directions on the topic of deep and graph learning methods for DDIs prediction as illustrated in Table 1, and we further investigate the comparison performance of these representative baseline models as shown in Tables 3–5.
(iii) *Abundant resources*. We have gathered a comprehensive collection of resources dedicated to DDIs prediction. These collections include open-sourced deep and graph learning methods, available platform and toolkit, as well as an important paper list. These resources can be accessed our github (https://github.com/xzenglab/resources-for-DDIs-prediction-using-DL), which will be continuously updated.
(iv) *Future directions*. We discuss the limitations of existing works and suggest several promising future directions.

DATA SOURCES

DDIs is not only known as a binary relationship but also can be affected by numerous factors, such as chemical substructures, targets and enzymes. The available datasets collect multiple drug-related information including mechanism of actions, protein structures and pharmacogenomic effects, which provide great opportunity for scientific communities to effectively develop novel methods to predict various drug interactions. In this section, we provide a brief overview of several commonly used chemical and bioinformatics databases for DDIs prediction as shown in Table 2.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

Table 1. List of deep and graph learning models for DDI prediction

Model	Year	Input	Representation	Architecture	Task	Code
DeepDDI [13]	2018	SMILES	Structural similarity	FC layers	Binary/multi-class classification	Link
Decagon [14]	2018	SMILES	Heterogeneous network	Encoder + decoder	Binary classification	Link
MLRDA [15]	2019	SMILES	Drug features	Encoder + decoder	Binary classification	-
KMR [16]	2019	Drug ID	Multiple drug descriptor	CNN, Bi-LSTM + attention	Binary classification	-
MR-GNN [17]	2019	SMILES	Molecular graph	Weighted GCN + LSTM	Binary/multi-class classification	Link
MHCADDI [18]	2019	SMILES	Molecular graph	GCN + co-attention	Binary/multi-label classification	Link
D^3I [19]	2019	Drug ID	Drug features	Encoder + aggregator	Binary classification	Link
KG-DDI [20]	2019	Drug ID	Knowledge graph	Conv-LSTM	Binary classification	Link
DDIMDL [21]	2020	Drug name	Diverse drug features	DNN	Multi-class classification	Link
GoGNN [22]	2020	SMILES	Molecular/interaction graph	GNN + attention	Multi-class/multi-label classification	Link
KGNN [23]	2020	Drug ID	Knowledge graph	GNN	Binary classification	Link
BioBERT [24]	2020	SMILES	Embedding	BERT	Binary classification	Link
SkipGNN [25]	2020	Drug ID	Skip graph	MPNN	Binary classification	Link
CASTER [26]	2020	SMILES	Substructure	Encoder + decoder	Binary classification	Link
EPGCN-DS [27]	2020	SMILES	Molecular graph	GCN	Binary classification	Link
HOGCN [28]	2021	Drug ID	Interaction network	High-order GCN	Binary classification	Link
MUFFIN [29]	2021	SMILES/Drug ID	Molecular graph + knowledge graph	MPNN + TransE	Binary/multi-class/multi-label classification	Link
SumGNN [30]	2021	SMILES/Drug ID	Knowledge graph/subgraph	GNN + attention	Multi-class/multi-label classification	Link
MIRACLE [31]	2021	SMILES	Molecular graph	GCN + Contrastive learning	Binary classification	Link
SSI-DDI [32]	2021	SMILES	Substructure	GAT + Co-attention	Binary classification	Link
AAEs [33]	2021	Drug ID	Knowledge graph	Adversarial autoencoders	Binary classification	Link
GNN-DDI [34]	2022	SMILES	Molecular graph	GAT	Binary classification	Link
MFFGNN [35]	2022	SMILES + molecular graph	Multi-type feature	GNN + BiGRU	Binary classification	Link
GCNMK [36]	2022	Drug ID	DDI graph + drug features	GCN + Linear transformation	Binary classification	-
DeepDrug [37]	2022	SMILES	Molecular graph	RGCN	Binary/multi-class/multi-label classification	Link
LR-GNN [38]	2022	Drug ID	Biomedical network	GCN	Binary classification	Link
DANN-DDI [39]	2022	Drug ID	Biomedical network	SDNE + attention	Binary classification	Link
DGAT-DDI [40]	2022	Directed graph	Source/target encoding	Source/target GAT	Binary classification	Link
GMPNN [41]	2022	SMILES	Molecular graph	Gated MPNN	Binary classification	Link
STNN-DDI [42]	2022	SMILES	Substructure	Encoder + decoder	Binary classification	Link
deepMDDI [43]	2022	Drug ID	Sub-networks	RGCN Encoder + decoder	Multi-label classification	Link
RANEDDI [44]	2022	Drug ID	DDI network	RotatE + network embedding	Binary/multi-class classification	Link
DeSIDE-DDI [45]	2022	Fingerprints	Gene expressions	DNN	Multi-class classification	Link
SA-DDI [46]	2022	SMILES	Substructure	D-MPNN	Binary classification	Link
MSAN [47]	2022	SMILES	Substructure	Transformer-like framework	Binary classification	Link
LaGAT [48]	2022	Drug ID	Knowledge graph/subgraph	Link-aware GAT	Binary/multi-class classification	Link
Molormer [49]	2022	2D structures	Molecular graph spatial structure	Attention + Siamese network	Binary classification	Link
MDDI-SCL [50]	2022	Drug ID	Drug features	Attention + Contrastive learning	Multi-class classification	Link
R^2-DDI [51]	2022	SMILES	Molecular graph	DeeperGCN + Feature refinement	Binary classification	Link
BioDKG-DDI [52]	2022	SMILES	Multiple drug features	Attention + DNN	Binary classification	-
AMDE [53]	2022	SMILES	Sequence + atomic graph	MPAN + Transformer	Binary classification	Link
DDKG [54]	2022	SMILES/Drug ID	Knowledge graph	Encoder-decoder + GCN	Binary classification	Link
3DGT-DDI [55]	2022	3D structures	Molecular graph + position information	3D GNN + text attention	Binary/multi-class classification	Link
DSN-DDI [56]	2023	Molecular graph	Substructure	Dual-view encoder + decoder	Binary classification	Link
DGNN-DDI [57]	2023	SMILES	Molecular graph + substructure	Directed MPNN + substructure attention	Multi-class classification	Link
KG2ECapsule [58]	2023	Drug ID	Knowledge graph	GCN + Capsule	Multi-label classification	Link

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

Table 2. The widely used databases for DDI prediction.

Database	Publication year	Num. of drug	Num. of drug-related pairs	Latest update	Link
KEGG [59]	1995	11 147	324 183 DDIs	V104.1, 2022-11-01	Link
DrugBank [60]	2006	1706	191 808 DDIs	V5.1.9, 2022-01-03	Link
SIDER [61]	2008	1430	139 756 drug-side effect pairs	V4.1, 2015-10-21	Link
TWOSIDES [62]	2012	645	4 649 441 DDIs	–	Link
OFFSIDES [62]	2012	1332	18 842 drug-event associations	–	Link
BIOSNAP [63]	2018	1332	41 520 DDIs	–	Link

Table 3. Performance evaluation under binary classification task

Method	Year	AUPRC	ACC	AUROC	F1	Remarks
Dataset 1: DrugBank[a]						
DeepWalk [64]	2014	0.9070	0.8349	0.9181	0.8357	Network based method
GreRep [65]	2015	0.9115	0.8443	0.9230	0.8461	Network based method
LINE [66]	2015	0.8915	0.8280	0.9092	0.8318	Network based method
SDNE [67]	2016	0.8782	0.8303	0.9029	0.8373	Network based method
GAE [68]	2016	0.7403	0.7491	0.8085	0.7889	Network based method
struc2vec [69]	2017	0.8672	0.7882	0.8735	0.7962	Network based method
KG-DDI [20]	2019	–	0.7867	0.7867	0.7843	Network based method
KGNN [23]	2020	0.9892	0.9561	0.9912	0.9566	Network based method
AAEs [33]	2021	0.7899	–	0.9480	–	Network based method
DANN-DDI [39]	2022	0.9709	**0.9962**	0.9763	0.9692	Network based method
DGAT-DDI [40]	2022	0.943	0.886	0.951	0.884	Network based method
RANEDDI [44]	2022	**0.9894**	–	0.9898	0.9562	Network based method
AMDE [53]	2022	–	0.9763	0.9901	0.9760	Network based method
DeepDDI [13]	2018	0.828	–	0.844	0.772	Chemical structure based method
KMR [16]	2019	0.9568	0.9219	0.9512	0.9191	Chemical structure based method
CASTER [26]	2020	0.829	–	0.861	0.796	Chemical structure based method
SSI-DDI [32]	2021	0.9814	0.9447	0.9838s	–	Chemical structure based method
MFFGNN [35]	2022	0.9681	–	0.9539	0.9254	Chemical structure based method
DeepDrug [37]	2022	0.98	–	–	0.94	Chemical structure based method
GMPNN [41]	2022	–	0.9530	0.9846	–	Chemical structure based method
SA-DDI [46]	2022	–	0.9623	0.9880	0.9629	Chemical structure based method
MSAN [47]	2022	–	0.9700	0.9927	0.9704	Chemical structure based method
R^2-DDI [51]	2022	–	0.9815	**0.9970**	**0.9816**	Chemical structure based method
3DGT-DDI [55]	2022	–	–	0.970	–	Chemical structure based method
DSN-DDI [56]	2023	–	0.9694	0.9947	0.9693	Chemical structure based method
MIRACLE [31]	2021	0.9234	–	0.9551	0.8360	Hybrid method
BioDKG-DDI [52]	2022	–	0.9370	0.9830	0.9390	Hybrid method
Dataset 2: TWOSIDES[b]						
MR-GNN [17]	2019	–	0.7623	0.85	0.7788	Chemical structure based method
MHCADDI [18]	2019	–	–	0.8820	–	Chemical structure based method
SSI-DDI [32]	2021	–	0.7820	0.8585	0.7981	Chemical structure based method
DeepDrug [37]	2021	–	–	–	0.84	Chemical structure based method
GMPNN [41]	2022	–	0.8283	0.9007	0.8408	Chemical structure based method
SA-DDI [46]	2022	–	0.8745	0.9317	0.8835	Chemical structure based method
R^2-DDI [51]	2022	–	0.8615	0.9149	0.8731	Chemical structure based method
DSN-DDI [56]	2023	–	**0.9883**	**0.9990**	**0.9883**	Chemical structure based method

[a]The performance on DrugBank dataset of DeepDDI was directed from CASTER results, and that of DeepWalk, GreRep, LINE, SDNE, GAE, struc2vec and KG-DDI were reported from KGNN results, and that of other methods were directly obtained from original papers. The division of the train and test set might be different for each model. [b]The performance on TWOSIDES dataset of MR-GNN, MHCADDI, SSI-DDI, GMPNN and SA-DDI were reported from DSN-DDI results, and that of other methods were directly obtained from original papers. The division of the train and test set might be different for each model. Bold values significance the greater these evaluation metrics the better the prediction.

KEGG

KEGG database is originally used to discover utilities of the biological system and high-level functions, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. As an integrated database with 16 resources, it was broadly classified into systems, genomic, chemical and health information, such as KEGG PATHWAY and KEGG DRUG. As for KEGG DRUG, it collects multiple drug information of approved drugs and unifies them according to their chemical structures. Specifically, each entry is identified by the drug number and associated with KEGG original annotations (e.g. drug metabolism), which results in 1925 approved drugs and their 56 983 interactions spanning 11 147 drugs and 324 183 interactions, respectively.

DrugBank

DrugBank is a free-to-access and online database that collects drugs, drug targets, their mechanisms and interactions. Version 1.0 started in 2006 and the latest version has been updated to 5.1.9 in 2022. At present, it contains 14 944 drug entries, including

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

Table 4. Performance evaluation under multi-class classification task

Method	Year	Mean accuracy	Macro precision	Macro recall	Macro F1	Remarks
Dataset 1: DrugBank[a]						
DeepWalk [64]	2014	0.8000	0.8220	0.7101	0.7469	Network based method
LINE [66]	2015	0.7506	0.6870	0.5451	0.5804	Network based method
Decagon [14]	2018	0.8719	–	–	0.5735	Network based method
KG-DDI [20]	2019	0.8923	0.7945	0.7667	0.7666	Network based method
KGNN [23]	2020	0.9127	0.8583	0.8170	0.8291	Network based method
SkipGNN [25]	2020	0.8583	–	–	0.5966	Network based method
SumGNN [30]	2021	0.9266	–	–	0.8685	Network based method
LaGAT [48]	2022	0.9604	–	–	0.9289	Network based method
DeepDDI [13]	2018	0.8371	0.7275	0.6611	0.6848	Chemical structure based method
DDIMDL [21]	2020	0.8852	0.8471	0.7182	0.7585	Chemical structure based method
SSI-DDI [32]	2021	0.8965	0.8763	0.9321	0.8993	Chemical structure based method
GMPNN [41]	2022	0.9485	0.9346	0.9725	0.9495	Chemical structure based method
SA-DDI [46]	2022	0.9565	0.9472	0.9746	0.9573	Chemical structure based method
Molormer [49]	2022	**0.9667**	0.9419	0.9270	0.9311	Chemical structure based method
MDDI-SCL [50]	2022	0.9378	0.8804	0.8767	0.8755	Chemical structure based method
DGNN-DDI [57]	2023	0.9609	0.9472	**0.9788**	**0.9616**	Chemical structure based method
MUFFIN [29]	2021	–	**0.9648**	0.9495		Hybrid method method

[a]The performance on DrugBank dataset of DeepWalk, LINE, DeepDDI, KG-DDI and KGNN were reported from MUFFIN results, and that of Decagon and SkipGNN were obtained from SumGNN results, and that of GMPNN, SA-DDI and SSI-DDI were obtained from DGNN-DDI results, and that of other methods were directly obtained from original papers. Bold values significance the greater these evaluation metrics the better the prediction.

Table 5. Performance evaluation under multi-label classification task

Method	Year	PR-AUC	ACC	ROC–AUC	F1	Remarks
Dataset 1: DrugBank[a]						
DeepWalk [64]	2014	0.4782	0.6163	0.6501	0.5861	Network based method
LINE [66]	2015	0.4923	0.6374	0.6926	0.6190	Network based method
KGNN [23]	2020	0.8587	0.7947	0.8602	0.7945	Network based method
KG2ECapsule [58]	2023	**0.8858**	**0.8050**	**0.8882**	**0.8145**	Network based method
Dataset 2: TWOSIDES[b]						
DeepWalk [64]	2014	0.6160	–	0.8708	–	Network based method
LINE [66]	2015	0.6043	–	0.8621	–	Network based method
node2vec [126]	2016	0.8887	–	0.9066	–	Network based method
Decagon [14]	2018	0.9060	–	0.9172	–	Network based method
KG-DDI [20]	2019	0.6527	–	0.8906	–	Network based method
KGNN [23]	2020	0.6584	–	0.8948	–	Network based method
SkipGNN [25]	2020	0.9090	–	0.9204	–	Network based method
SumGNN [30]	2021	**0.9335**	–	**0.9486**	–	Network based method
DeepDDI [13]	2018	0.5032	–	0.8301	–	Chemical structure based method
MUFFIN [29]	2021	0.7033	–	0.9160	-	Hybrid method

[a]The performance on DrugBank dataset of DeepWalk, LINE and KGNN were reported from KG2ECapsule results. [b]The performance on TWOSIDES dataset of DeepWalk, LINE, node2vec, Decagon, KG-DDI, KGNN and SkipGNN were reported from SumGNN results, and that of DeepDDI was obtained from MUFFIN results, and that of other methods were directly obtained from original papers. Bold values significance the greater these evaluation metrics the better the prediction.

2729 approved small molecule drugs, 1564 approved biologics (e.g. proteins and allergenics) and over 6713 experimental drugs including discovery-phase. Generally, given two drugs with their SMILES sequences, the final goal is to predict their interaction type (i.e. binary, multi-class and multi-label classification). DrugBank V5.1.4 is widely used in comparison experiment, and it contains 1706 drugs and 191 808 drug pairs with 86 DDI types.

SIDER

Side effect resource collects multiple information from marketed drugs and their side effects to provide a more comprehensive view of actions of drugs and their adverse reactions. It can predict the potential side effects of drug candidates according to their binding fingerprints, chemical structures and other chemical properties. Meanwhile, it combines side effect information with other resources in chemical biology, which will greatly benefit pharmacology and medical research. Its current version 4.1 includes 1430 drugs, 5868 side effects and 139 756 drug-side effect pairs.

TWOSIDES

The TWOSIDES databases collect polypharmacy side effects that are related to individual one in the drug pairs or higher-order drug combinations. Overall, it contains 868 221 associations between 59 220 pairs of drugs and 1301 adverse events. Additionally, it contains 3 782 910 significant associations for which the drug pair has a higher side-effect association score, evaluated by the proportional reporting ratio (PRR) [70], than those of the individual drugs alone. Specifically, it contains 645 drugs and side effects caused by 63 473 combinations of different drugs. Generally, given two drugs with their SMILES sequences, the final goal is to predict all side effects (i.e. multi-label classification).

OFFSIDES

The OFFSIDES database collects 438 801 off-label side effects between 1332 drugs and 10 097 adverse events. Off-label means no record on the US Food and Drug Administration (FDA)'s official drug label while on-label means the opposite. The drug label lists

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

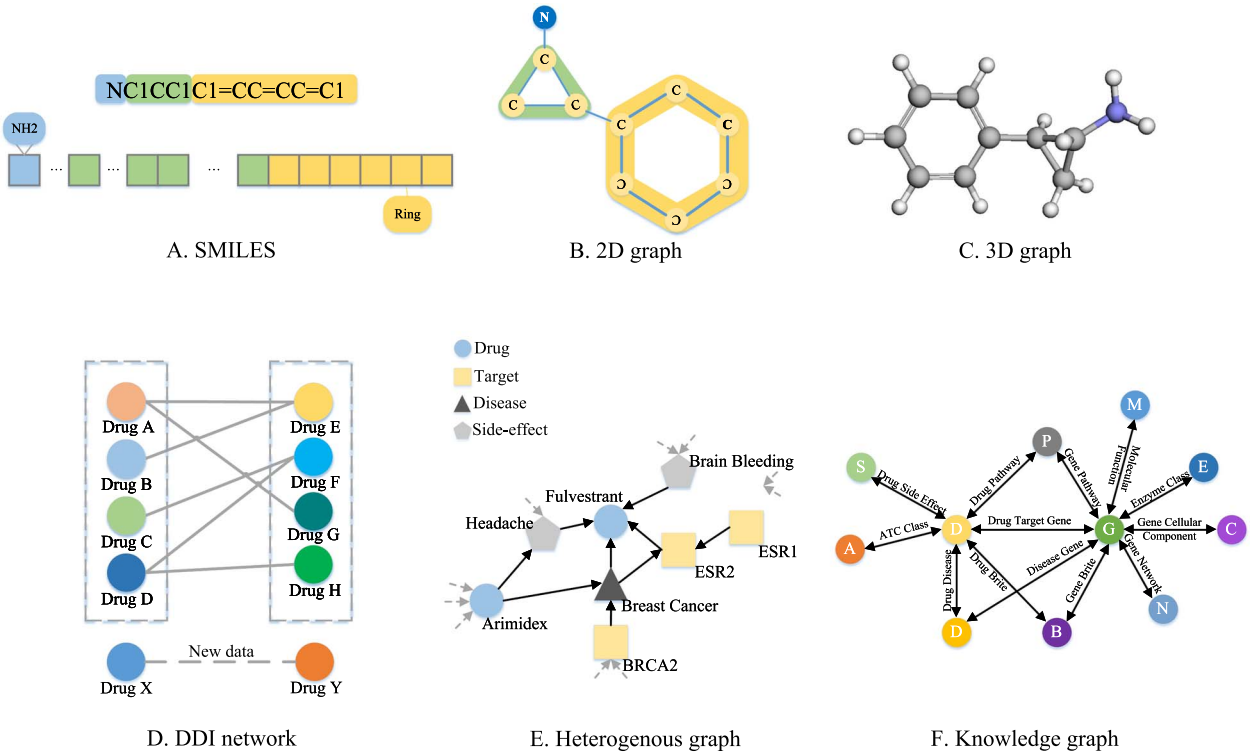

Figure 2. A diagram illustrating six commonly-used molecular representation approaches, including: (A) one-dimensional (1D) sequence-based representation; (B) 2D graph-based representation; (C) 3D representation; (D) DDI network; (E) heterogeneous graph and (F) knowledge graph.

in average 69 on-label adverse events. And it listed an average of 329 high-confidence off-label adverse events for each drug. Moreover, it recovers 38.8% (i.e. 18 842 drug-event associations) of SIDER associations from the adverse event reports.

BIOSNAP

The BIOSNAP dataset collects various types of interactions between FDA-approved drugs by constructing a biological network. Nodes represent drugs and edges represent drug interactions. This dataset contains 1322 approved drugs with 41 520 labeled DDIs that are extracted from drug labels and scientific publications.

MOLECULAR REPRESENTATION

The representation of drug molecule is an crucial part in drug-related tasks, including DDIs prediction. For example, Tranylcypromine is an inhibitor of the enzyme monoamine oxidase [71], functioning nonselectively and irreversibly, and thus it is also employed clinically as an antidepressant and anxiolytic agent in the treatment of mood and anxiety disorders. Take its SMILES (Simplified Molecular Input Line Entry System) [72] format *C1C(C1N)C2=CC=CC=C2* as an example, Tranylcypromine is represented by six commonly used molecular representation as shown in Figure 2.

Sequences

As the most frequently used molecule descriptor, SMILES is a string of characters as shown in Figure 2A, where each atom is encoded by a respective ASCII symbol, and chemical bonds, branching as well as stereochemistry are represented by specific symbols in SMILES strings. The SMILES sequence is capable of converting the chemical structure into a spanning tree by utilizing a longitudinal-first traversal tree algorithm to generate a sequence of characters. A variety of deep learning models, such as recurrent neural networks, are able to employ their internal state (memory) to process variable length sequences of inputs [73–75], using SMILES sequences as input to extract the chemical context via various natural language processing techniques, including Mol2Vec [76] and FCS [77]. Sequence-based representations tend to be compact, memory-efficient and easily searchable.

2D graph

A more direct way to representing drug molecules is through 2D graph-based representation (i.e. molecular graph) as shown in Figure 2B. In particular, we denote 2D graph as $G_{2D} = (X, E)$, $X \in \mathbb{R}^{N \times d}$ represents the atom attribute matrix, where N denotes the number of nodes and d denotes the dimensionality of node feature, and E are characterized by the type of chemical bonds between the atoms, including single, double, triple and aromatic bond. Specifically, Figure 3 shows an example of the molecular graph representation of *Tranylcypromine*. First, the SMILES sequence is transformed into its 2D structures using RDKit tool. Predefined atomic features are then assigned to each node based on its atom number. In a molecular graph, each node contains a 78-dimension initial feature vector to encode five types of atomic features, including atomic symbol, adjacent atoms, adjacent hydrogens, implicit value and aromaticity. Finally, we obtain the molecular graph representation of *Tranylcypromine* that consists of atom number (i.e. total number of atoms), atomic features and edge features (i.e. edge list). This representation allows us to extract the structural information from a molecular graph. We then typically apply a transformation function T_{2D} to the topological graph. Given a 2D graph G_{2D} or molecular graph obtained from its SMILES sequence via RDKit [78], its

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

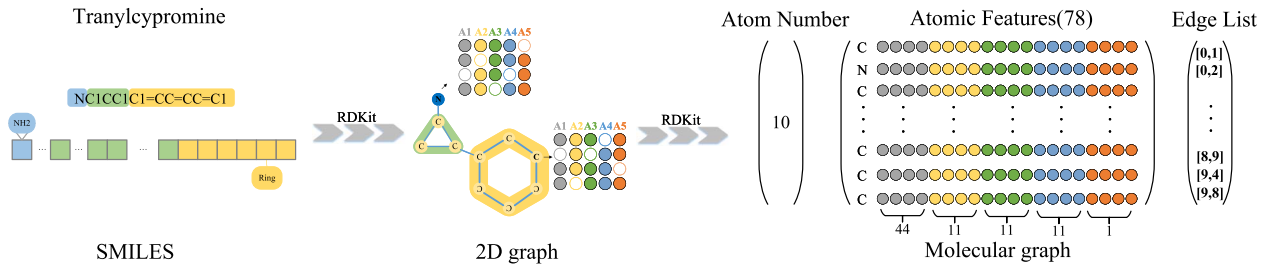

Figure 3. An example of the molecular graph representation of *Tranylcypromine* by using RDKit.

representation H_{2D} can be computed from a 2D GNN model:

$$H_{2D} = 2DGNN(T_{2D}(X, E)). \quad (1)$$

Usually, message passing neural networks (MPNN) [79] known as one of the classic 2D GNN models are designed to accomplish the encoding of graph-based methods. Since 2D graph is usually stored in the form of adjacency matrices. The utilization of 2D GNN not only allows for faster and accurate combination of properties between two adjacent atoms or chemical bonds, but it also allows the weights to be optimized in the message passing process. With comparison to sequence-based approaches, graph-based representations are easy to extract the structural information via graph convolutional operations, where bond weights can be updated and optimized in message-passing networks.

3D graph

Many methods use sequence- and graph-based representations of molecules as the inputs, such as SMILES sequences and molecular graphs. Although these methods can effectively preserve the structural information of drug molecules, they can not capture well the inter-binding relationships between ligands and receptors, especially biologically meaningful in 3D relationships, because the 3D coordinates of all atoms in the ground-state molecule are critical to various applications, including molecular property prediction [80] and molecular conformer ensembles [81]. Here, 3D graph as shown in Figure 2C represents the spatial arrangements of each atom in the 3D space, containing a list of atoms with atom types and atomic coordinates. Generally, each molecule with n atoms is expressed as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^{n}$ is the set of vertices symbolizing atoms and $\mathcal{E} = \{e_{ij} | (i,j) \in |\mathcal{V}| \times |\mathcal{V}|\}$ is the set of edges representing interatomic bonds. Each node $v_i \in \mathcal{V}$ indicates the atomic attributes (e.g. the element type), and each edge $e_{ij} \in \mathcal{E}$ describes the connection between v_i and v_j, and is labeled with its chemical type. Additionally, we also assign virtual types to the unconnected edges. For 3D geometry graph, each atom in $\mathcal{V}$ is embedded in the 3D space with a coordinate vector $\mathbf{c} \in \mathbb{R}^3$, and the full set of positions (i.e. the conformation where atoms are represented as their Cartesian coordinates) can be represented as $C = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n]$, where $\mathbf{c}_i \in \mathbb{R}^3$. Then we generally apply a transformation function T_{3D} on the geometry graph. Given a geometry graph $G_{3D} = (X, C)$, its representation H_{3D} can be obtained via a 3D GNN model:

$$H_{3D} = 3DGNN(T_{3D}(X, C)). \quad (2)$$

DDI network

DDIs can be associated with biological, chemical and phenotypic information about drugs. The drug–drug interaction network (DDI network) is proposed to learn the potential associations between drug molecules. Generally, the DDIs prediction problem is formulated as a missing link prediction task by constructing a DDI network with drugs as nodes and known interactions as edges. Specifically, drugs should be represented as feature vectors via interaction profile from known interactions to build prediction models. As shown in Figure 2D, let *A, B, C,..., H* be a set of given drugs, the drug interaction profile, which is a binary vector indicating the presence or absence of interaction between drugs (e.g. $A \longleftrightarrow E$, $A \longleftrightarrow G$), can be represented as an interaction network.

Heterogeneous graph

Heterogeneous graph (HetG) contains a wealth of information with structural relations (i.e. edges) among nodes of various types, as well as unstructured content associated with each node [82]. For example, HetG can be expressed as to involve many other types of biological entity relationships in the process of predicting DDIs. Considering these different associations can enhance the prediction performance. In general, the HetG associated with DDIs is expressed as a graph $G = (V, E, O_V, R_E)$, where V and E denote the sets of nodes and links, respectively. O_V and R_E represent the set of object types and that of relation types, respectively. Furthermore, each node is associated with heterogeneous contents (e.g. attributes). Specifically, the HetG denotes relations between different pairs, including drugs and targets, drugs and side-effects, drugs and diseases. For instance, Figure 2E illustrates the biological heterogeneous graph centered around drug *Fulvestrant*, where edges with distinct colors denote different relations, and arrows indicate the direction of information flow.

Knowledge graph

Recently, knowledge graphs (KGs), which are a form of structured human knowledge, have been gaining increasing attention from both academic and various aspects of the drug discovery domain [83]. For example, KGs can be utilized to better integrate multiple entity types and diverse association relations between biological entities. This approach allows for the extraction of high-order semantic features to improve DDIs prediction. KGs can be represented by a structured representation of facts, which comprised of entities, relationships and semantic descriptions. Generally, a knowledge graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$, where $\mathcal{E}, \mathcal{R}$ and $\mathcal{F}$ are sets of entities, relations and facts, respectively. A fact is denoted as a triple $(h, r, t) \in \mathcal{F}$. As shown in Figure 2(F), entities (i.e. nodes) with different color and alphabet represent real-world biological objects (e.g. drug, target and side-effect), and relationships (i.e. edges) depict the connection between entities, where semantic descriptions of entities and their relationships encompass types and properties with a clearly defined meaning, including *Drug Disease*, *Drug Target Gene* and *Drug Brite*. As a topical concrete application, KGs have been utilized in helping to combat the COVID-19 pandemic [84, 85]. Additionally, there are few existing knowledge graph covering various aspects of the drug discovery process, including Hetionet [86], DRKG [87], BioKG [88],

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

PharmKG [89], OpenBioLink [90] and Clinical Knowledge Graph [91]. Note that a comprehensive review is beyond the scope of this work and readers who are interested are directed to a dedicated review [92].

MODELS

The DDIs prediction is to develop a computational model that receives two drugs with an interaction type as inputs and generates an output prediction indicating whether there exists an interaction between them. As introduced in previous section (i.e. Molecular Representation), drug molecules generally use RDKit to convert its SMILES sequence into molecular graphs with nodes as atoms and edges as chemical bonds. Specifically, a graph for a given SMILES is denoted by $G = (V, E)$, where V is the set of N nodes represented by a d-dimensional vector, and E is the set of edges represented as an adjacency matrix A. In a molecule graph, x_i (resp., x_j) $\in V$ is the i (resp., j)th atom and $e_{ij} \in E$ is the chemical bond between the x_i and x_j. Owing to the non-Euclidean and translation invariance, GNNs have been proposed to replace traditional convolution networks in order to extract drug feature representations from chemical molecular graph. In the case of GNN, the process of learning drug representation is essentially the message passing between each node and its neighboring nodes. Thus we further systematically review four types of GNN models for encoding molecular representations into continuous vectors, as shown in Figure 3, including graph convolutional network, message passing neural network, graph attention network and graph auto-encoder.

Graph convolutional networks

Graph convolutional networks (GCNs) are a class of neural networks specifically designed for graph-structured datasets. Advances in this direction are often categorized as spectral- and spatial-based approaches. Spectral-based approaches learn compact representations of graph elements, their attributes and supervised labels as shown in Figure 4A. Following the original paper of GCN [93], the input of multi-layer GCN is the node feature matrix $X \in \mathbb{R}^{N\times d}$ and the adjacency matrix $A \in \mathbb{R}^{N\times N}$ that represents the connection of nodes. The layer-wise propagation algorithm can be obtained as below:

$$H^{l+1} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}), \tag{3}$$

where $\tilde{A} = A + I_N$ is the adjacency matrix of an undirected graph $\mathcal{G}$ with added self-connections, and I_N represents the identity matrix, $\tilde{D}$ is diagonal matrix with $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ and $W^{(l)}$ is a layer-specific trainable weight matrix. Here, $\sigma(\cdot)$ denotes an activation function (e.g. $ReLU(\cdot) = max(0, \cdot)$). And $H^{(l)}$ (resp., $H^{(l+1)}$) $\in \mathbb{R}^{N\times D}$ represents the matrix of activations in the l (resp., $l + 1$)th layer, respectively. We suppose that $H^{(0)} = X$. The output $Z \in \mathbb{R}^{N\times F}$ (F is the number of output features for every node) can be obtained as below:

$$Z = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}X\Theta), \tag{4}$$

where $\Theta \in \mathbb{R}^{F\times d}$ represents the matrix of filter parameters.

In contrast, spatial-based approaches [94] define convolutions directly on the graph by propagating and aggregating node representations from neighboring nodes in the vertex domain, as opposed to spectral-based GCN, which depends on the specific eigenfunctions of the Laplacian matrix. Following KGNN [23], the proposed method learns latent representations of drugs and their neighborhood entities embedding between drug pairs from the constructed KG. Figure 4A shows an example of a two-layer KGNN of the given G node (green) in a KG. Note that besides the immediate neighbors (e.g. B, E and P), it also extends KGNN to two-layer ($H = 2$) to extract both high-order structures and semantic relations. Generally, given a node v at the kth depth and its graph convolution is computed by:

$$\mathbf{h}^k_{\mathcal{N}(v)} \leftarrow \text{AGGREGATE}_k\left(\{\mathbf{h}^{k-1}_u, \forall u \in \mathcal{N}(v)\}\right), \tag{5}$$

$$\mathbf{h}^k_v \leftarrow \sigma\left(W^k \cdot \text{CONCAT}(\mathbf{h}^{k-1}_v, \mathbf{h}^k_{\mathcal{N}(u)})\right), \tag{6}$$

where each node $v \in V$ aggregates the representation vectors of all its immediate neighboring nodes $u \in \mathcal{N}(v)$ in the current depth via some learnable AGGREGATE operation. Then it combines the node's current representation $\mathbf{h}^{k-1}_v$ with its aggregated neighborhood representation $\mathbf{h}^{k-1}_{\mathcal{N}(v)}$, and finally passes the combined vector to a fully-connected layer with a nonlinear activation function $\sigma(\cdot)$, followed by a normalization step. And the output of final representation at depth K are denoted by $\mathbf{z}_v = \mathbf{h}^K_v$.

The aggregator functions include *mean*, *LSTM* and *pooling* aggregators. The *mean* aggregator can be simplified as follows:

$$\mathbf{h}^k_v \leftarrow \sigma\left(W \cdot \text{MEAN}(\{\mathbf{h}^{k-1}_v\} \cup \{\mathbf{h}^{k-1}_u, \forall u \in \mathcal{N}(v)\})\right), \tag{7}$$

$$\text{AGGREGATE}^{pooling}_k = \max\left(\{\sigma(W_{pool}\mathbf{h}^k_{u_i} + b), \forall u_i \in \mathcal{N}(v)\}\right). \tag{8}$$

Message passing neural network

Message passing neural network (MPNN) is a typical type of GNNs that maps an undirected graph $\mathcal{G}$ to a graph-level vector $h_{\mathcal{G}}$ using *Message passing* and *readout*. As depicted in Figure 4B, *message passing* is first used to update node-level features (i.e. V_0) by aggregating messages from their neighbor nodes (i.e. V_1, V_2 and V_3). Following that, the *readout* process is designed to generate a graph-level feature vector by aggregating all the node-level features from a molecule graph. Finally, a label is predicted for the graph based on the graph-level feature vector. Concretely, the *Message passing* consists of T steps. On each step t, node-level hidden feature $h^{(t)}_i$ and messages $m^{(t)}_i$ associated with each node v_i are updated using message function M_t and node update function U_t. Their definitions are as follows.

$$m^{t+1}_i = \sum_{v_j \in N(v_i)} M_t(\mathbf{h}^{(t)}_i, \mathbf{h}^{(t)}_j, e_{ij}), \tag{9}$$

$$\mathbf{h}^{t+1}_i = U_t(\mathbf{h}^{(t)}_i, m^{t+1}_i), \tag{10}$$

where $N(v_i)$ represents the set of neighbors of v_i in the graph $\mathcal{G}$, and $h^{(0)}_i$ is set to the initial atom features x_i. The *readout* then uses a readout function R to obtain a graph-level feature vector based on the node-level features at the final step as follows.

$$\mathbf{h}_g = R(\{\mathbf{h}^{(T)}_i | v_i \in \mathcal{G}\}). \tag{11}$$

The message function M_t, node update function U_t, and readout function R are all learned differentiable functions.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

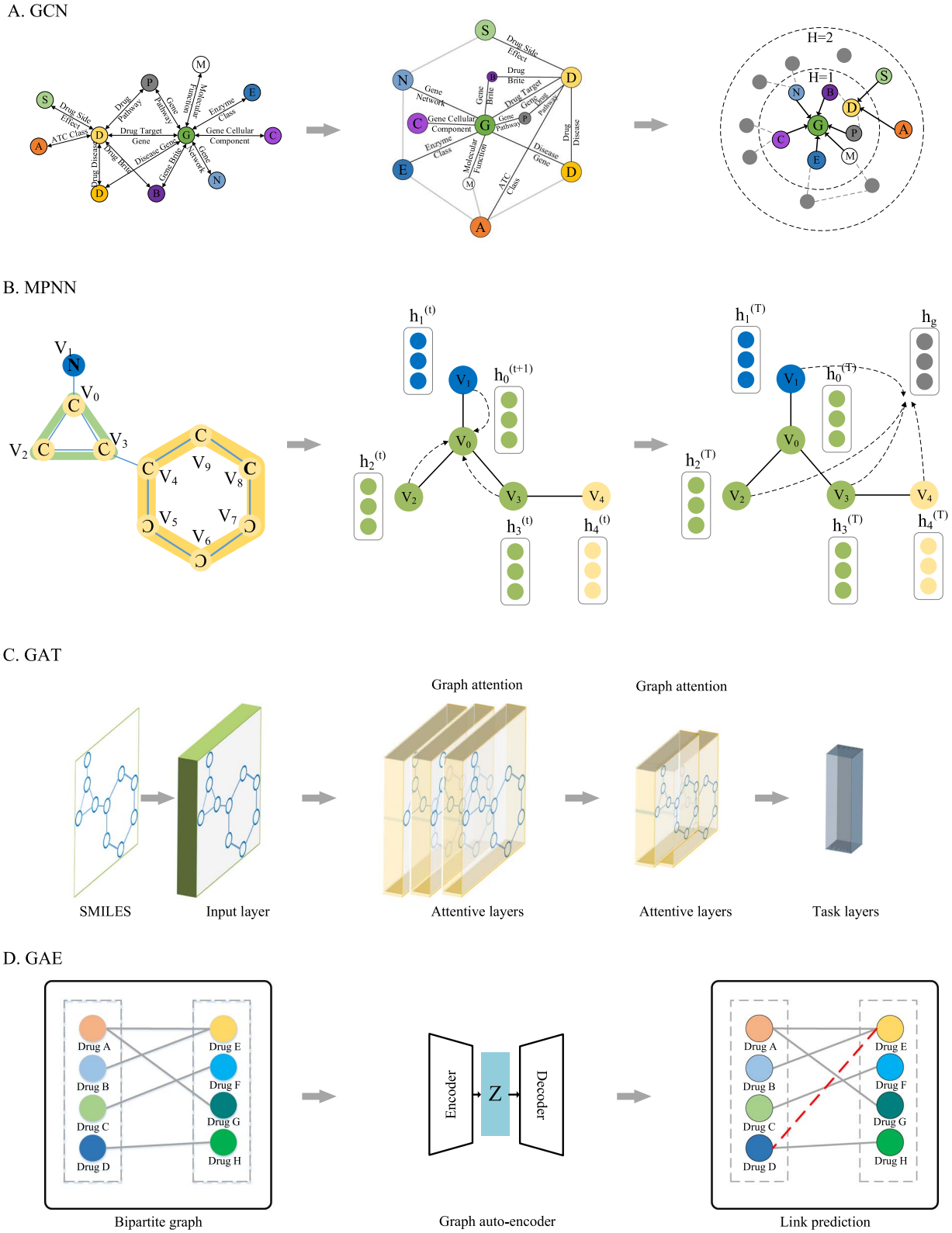

Figure 4. A diagram illustrating the theory framework of widely-used GNN models in the DDIs prediction, including: (A) GCN; (B) MPNN; (C) GAT; (D) GAE.

Graph attention network

Traditionally, GCN assigns the same weight to each neighbor node, and not every neighbor node has the same importance. Thus, graph attention network (GAT) is proposed to introduce a graph convolution model based on self-attention mechanism, which incorporates a *graph attention layer* in its architecture as shown in Figure 4C. According to the original paper of GAT [95], a set of node features $x \in R^F$ is used as input of GAT layer, and a linear

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

transformation is applied to each node based on a weight matrix $W \in R^{F \times F'}$, where F and F' are the dimensions of the input and output nodes, respectively. Moreover, *attention coefficients* between a node and its one-hop neighbors are adopted to obtain the output node as follows:

$$e_{ij} = \alpha(W\vec{h}_i, W\vec{h}_j), \tag{12}$$

where e_{ij} represents the importance of node j to node i. To ensure that the coefficients are comparable across different nodes, they are normalized across all choices of j using the softmax function as follows:

$$\alpha_{ij} = softmax_j(e_{ij}), \tag{13}$$

the non-linearity function σ is finally applied to compute the output node $\vec{h}_i'$ by:

$$\vec{\mathbf{h}}_i' = \sigma\left(\sum_{j \in N_i} a_{ij} W \vec{h}_j\right), \tag{14}$$

while a basic operation of *attention* is multi-head. Simply, it is to repeat the previous operation multiple times, but the parameters that need to be trained are different each time, so that we can extract more information. The process of multi-head can be computed by:

$$\vec{\mathbf{h}}_i' = \prod_{k=1}^{K} \sigma\left(\sum_{j \in N_i} a_{ij}^k W^k \vec{\mathbf{h}}_j\right), \tag{15}$$

where K is the number of heads.

Graph auto-encoder

Graph auto-encoder (GAE) has been widely used in the field of unsupervised learning on graph-structure data. Obtaining the suitable embeddings to represent nodes in the graph is not trivial, GAE adopt the encoder-decoder structure to realize the goal and to apply to the downstream tasks, such as link prediction. If we view drugs as nodes and DDI as links in a graph, DDIs prediction can be considered as a task to complete a DDI adjacency matrix. As shown in Figure 4D, the encoder represents drugs into scalars and decoders use these scalars to rebuild the whole graph by predicting the existence of a link between a pair of nodes/drugs. The encoder can be viewed as representation methods and decoder can be viewed as classifiers. Generally, GAE employs GCN as an encoder to obtain latent representations or embedding of nodes. This process can be expressed as follows:

$$Z = \mathbf{GCN}(X, A), \tag{16}$$

where Z represents the latent representations of all nodes, X and A represent the feature matrix of the node and adjacency matrix, respectively. Here X and A as input are then fed into GCN function, and we have:

$$\mathbf{GCN}(X, A) = \tilde{A}ReLU(\tilde{A}XW_0)W_1, \tag{17}$$

where $\tilde{A} = D^{-1/2}AD^{-1/2}$, W_0 and W_1 represent parameters to be learned. In short, GCN is equivalent to a function that takes node features and adjacency matrix as input and outputs node embedding. After that, GAE uses the inner-product as a decoder to reconstruct the original graph, the computation of reconstructed adjacency matrix $\hat{A}$ can be formulated by:

$$\hat{A} = \sigma(ZZ^T), \tag{18}$$

in order for the reconstructed adjacency matrix to be as close to the original adjacency matrix as possible. Because the adjacency matrix determines the structure of the graph.

PREDICTION TASKS

Due to the increasing amount of data and advanced algorithms, DL has led to breakthroughs in various domains [96–99], including in the application of DDIs and drug-related prediction tasks [100–104]. Figure 5 shows an illustrative pipeline of several DL methods.

In the beginning, this line of work develop effective representation method (see **Section 3**) to capture high-level hidden embeddings from various public datasets (see **Section 2**). Different from traditional machine learning based methods that heavily rely on the handcraft feature and domain knowledge, these approaches can learn more abstract information via deep architectures (see **Section 4**) without manually selecting and tuning features [105, 106], and the learned latent embeddings are finally used to predict on downstream tasks. There are many different types of classification tasks that may be encountered in DDIs prediction and specialized approaches to modeling that may be used for each, including binary, multi-class and multi-label classification.

Classification based predictive modeling involves assigning a class label to input sample. *Binary classification* refers to predicting whether interactions exist without determining their specific type, and *multi-class classification* involves predicting the specific type of DDIs between drug pairs. Following [29] in the model training, we generally optimized the model parameters by minimizing the cross-entropy loss in the binary and multi-label classification tasks, as described below:

$$\mathcal{L}_1 = -[y_{ij} log\hat{y}_{ij} + (1 - y_{ij}) log(1 - \hat{y}_{ij})], \tag{19}$$

where $\hat{y}_{ij}$ denotes the interaction label for drug pair (d_i, d_j) in binary classification task, and in multi-label task, each element y_{ij} is the one-hot vector with 86 elements (e.g. 86 DDI types in DrugBank dataset).

Multi-label classification involves predicting one or more DDIs type for each drug pair, the loss is defined as follows:

$$\mathcal{L}_2 = -\sum_{c=1}^{N_c} y_c log\hat{y}_c, \tag{20}$$

where N_c is the number of multi-class DDI types, $y_c \in 0, 1$ describes whether current type c is the same as the true label of sample pair, and $\hat{y}_c$ indicates the probability that the observed sample (d_i, d_j) belongs to type c.

PROGRESS AND TAXONOMY OF COMPUTATIONAL APPROACHES

Computational approaches mainly design effective algorithms to discover patterns by using public datasets retrieved from clinical texts [107], electronic health records [108, 109], and social media [110]. These methods can be roughly divided into chemical structure, network based, NLP based and hybrid methods. A Taxonomy of the different methods is shown in Figure 1. Furthermore, we

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

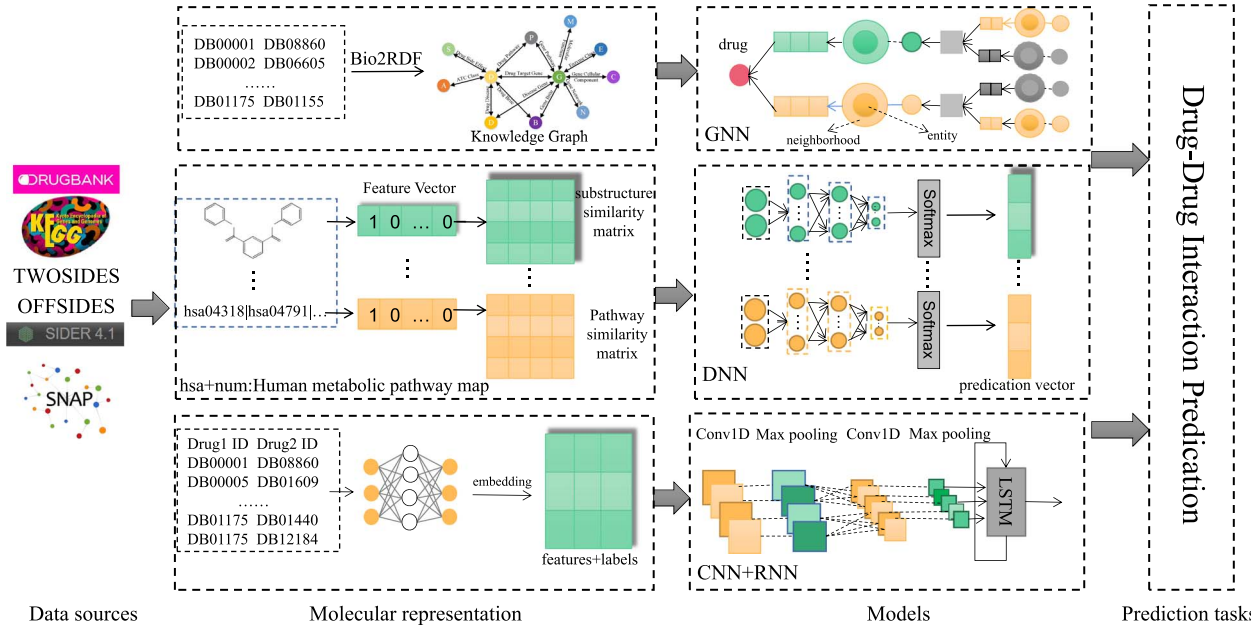

Figure 5. The pipeline of deep and graph learning methods for DDIs prediction. In general, drugs and their relevant information such as ID, name, and SMILES sequences are obtained from accessible *data sources* like DrugBank and TWOSIDES dataset. These drugs can then be optionally encoded into feature vectors using various *molecular representation* methods, such as molecular graph representation. The resulting representations, such as similarity matrices or 2D graphs, are subsequently fed into suitable *models*, such as GNNs, to generate interaction results or predicted scores based on the specific *prediction tasks*.

summarize and formulate 46 state-of-the-art deep and graph learning models in the recent years using a unified symbolic system in Table 1.

Chemical structure based methods

The vast majority of chemical structure based methods rely on similarity based, molecular graph and substructure based approaches, respectively.

Similarity based

Theses methods are based on the assumption that similar drugs may perform similar DDIs. They first extract some similarities from molecular structures [111], and various properties (e.g. phenotypic [112], functionality [113] and side effects [114]) as features for model training. Then they adopt classifiers to predict potential DDIs. For example, DeepDDI [13], which consists of structural similarity profile generation pipeline and deep neural network (DNN), is proposed to use the structural information to classify 86 DDIs types. MLRDA [15] is proposed to effectively exploit multiple drug features by leveraging a novel unsupervised disentangling loss CuXCov. Similarly, a knowledge-oriented DNN model is developed by KMR [16] to discover the interaction information among multiple features. Furthermore, D^3I [19] is presented to conduct cardinality- and order-invariant high-order DDIs prediction. DDIMDL [21] is constructed by the similarity assumption, and it built a multi-modal DL framework with multiple drug features to predict DDI events. A novel DL-based framework named DeSIDE-DDI [45] is developed to show more concern in interpretation on underlying genes, and it leveraged drug-induced gene expression signatures to engineer dynamic drug features by using a gating mechanism. Recently, a multi-type DDI prediction model named MDDI-SCL [50] is presented by supervised contrastive learning and three-level loss functions.

Molecular graph

Recent advances in artificial intelligence and technologies provide a set of potentially promising GNNs-based approaches for drug-related prediction tasks, including molecular property [115] and molecular interactions [116]. Naturally, drug molecules can be encoded by graph with atoms as nodes and chemical bonds as edges. Graph convolution neural networks (GCNs) have been proposed to extract node-level or graph-level features in various constructed graph [117]. For example, MR-GNN [17] is proposed to use multiple graph convolution layers to extract node features from different neighboring nodes in a structured entity graph. Moreover, MHCADDI [18] leverages a co-attentional mechanism to combine the type of side-effect and the molecular structures to obtain drug-level representation. EPGCN-DS [27] adopts a GCN based framework for type-specific DDI identification from molecular structures. GNN-DDI [34] learns k-hops drug representations its molecular graph via a five-layer GAT encoder. MFFGNN [35] combines the topological structure in molecular graphs with the interaction relationship between drugs and the local chemical context in SMILES sequences. Furthermore, Molormer [49] takes the 2D structures of drugs as input and encodes the molecular graph with spatial information based on a lightweight attention mechanism. DeepDrug [37] captures the intrinsic structural information of a compound by utilizing relational GCN module. Recently, R^2-DDI [51] further learns the drug representation by designing a relation-aware feature refinement framework.

Substructure based

Different from the aforementioned methods (i.e. MR-GNN) that takes the whole chemical structures into account, more recent efforts have attempted to leverage GNN for powerful feature extraction of drug substructures. A chemical substructure representation framework named CASTER [26] encodes the functional substructures of drugs. SSI-DDI [32] operates directly

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

on the raw molecular graph representations to identify pairwise interactions between their corresponding substructures. A gated MPNN (GMPNN) [41] learns chemical substructures with different sizes and shapes from the molecular graph representations. A substructure-aware tensor model, referred as to STNN-DDI [42], learns a 3D tensor to characterize a substructure-substructure interaction space. SA-DDI [46] develops a directed MPNN with attention mechanism to extract the size- and shape-adaptive substructures. A Transformer-like framework (MSAN) [47] extracts substructures via attention mechanism to associate atoms with learnable pattern vectors. DSN-DDI [56] employs local and global representation learning modules iteratively, and learns drug substructures from intra-view and inter-view simultaneously. DGNN-DDI [57] exploits the molecular structure and interaction information between chemical substructure via a co-attention mechanism. Furthermore, incorporating geometric information into GNNs to benefit some molecular prediction tasks has recently gained research attention [118], and 3D structures of drug molecules also contribute to DDIs tasks, where 3DGT-DDI [55] adopts 3D structural information of molecular graph and position information to improve the model performance, which can deeply explore the effect of drug substructure on DDI relationship.

Network-based methods

In general, network-based approaches infer the novel DDIs via label propagation [119, 120], multiple sources [121] or newly calculated features [122]. With the increasing availability of large biomedical network and the rapid development of deep learning, some studies attempt to incorporate with various advanced techniques, including graph embedding, link prediction and knowledge graph.

Graph embedding

Various graph embedding algorithms have been proposed to acquire potentially effective network-based features, including matrix factorization-based methods (e.g. GraRep [65], HOPE [123], FGRMF [124], BRSNMF [125]) that utilize the adjacency matrix as the input to learn latent embeddings from matrix factorization and random walk-based methods (e.g. DeepWalk [64], node2vec [126] and struc2vec [69]) that first generate sequences of nodes through random walks and then feed the sequences into the model to learn node representations, and neural network-based methods (e.g. Line [66], SDNE [67] and GAE [68]) that adopt different neural architectures and use various graph information as input. More detailed introduction of graph embedding on biomedical networks refer to [127]. Recently, GCNMK [36] obtains the embeddings of drugs by constructing two DDI graphs as the graph kernels.

Link prediction

Meanwhile, some GNN approaches cast the prediction as a link prediction problem on DDI graph or network. Previously, Decagon [14] is presented to develop a graph auto-encoder approach for multirelational link prediction on a multi-modal graph that consists of multiple interactions (e.g. drug–protein target interactions). In addition to aggregate information from direct interactions in biological network, a skip similarity approach named SkipGNN [25] that receives neural messages from two-hop neighbors and direct neighbors in the interaction network. Analogously, DANN-DDI [39] builds multiple drug feature networks and learns drug representations from these networks by using the graph embedding method. HOGCN [28] is introduced to adopt a higher-order GCN to gather different features from the higher-order neighborhood for biomedical interaction prediction. A GNN based on graph structure and initial features, named LR-GNN [38], constructs the link representation by designing a propagation algorithm to capture the node embedding. Furthermore, DGAT-DDI [40] is the first approach for predicting asymmetric interactions among drugs, and it designs a directed GAT to learn the embeddings of the source and the target role. deepMDDI [43] learns the topological features of DDI network by combining RGCN encoder with similarity regularization of multiple drug features. Recently, a relation-aware network embedding model, abbreviated RANEDDI [44], extracts the multirelational information and relation-aware network structure information together.

Knowledge graph

Existing GNN approaches for DDIs typically depend on one source of information, while using information from multiple sources could help improve predictions [54, 128]. Particularly, knowledge graph (KG) has greatly stimulated research on various domains, including relation inference and recommendation [129]. In our knowledge, KG-DDI [20] is the first specialized for DDIs task that embeds the nodes in the constructed KG using various embedding approaches. Another KG embedding framework (AAEs) [33] uses adversarial autoencoders based on Wasserstein distances and Gumbel-Softmax relaxation. Furthermore, KGNN [23] successfully adopts GCNs with neighborhood sampling to explicitly extract the neighborhood relations. More recently, subgraph structures have been found to contain rich information for many graph learning tasks. SumGNN [30] further uses KG to extract tractable pathway by designing a graph summarization module on subgraphs. And a link-aware graph attention method called LaGAT [48] generates multiple attention pathways for drug entities based on various drug pair links in KG. DDKG [54] further learns the drug embeddings from their attributes in the KG, and then simultaneously considers both neighboring node embeddings and triple facts by attention mechanism. KG2ECapsule [58] integrates capsule network to explicitly model the multi-relational DDI data based on biomedical KG.

NLP-based methods

As training DNNs from scratch often requires a large number of labeled data, which are expensive to acquire in real-world scenarios, inspired by the recent success in NLP, pre-trained models have been proposed to learn universal molecular representations from massive unlabeled molecules and fine-tuned on downstream tasks with task-specific labeled data. BioBERT [24] is first introduced to investigate how the pre-trained language model BERT [130] are pre-trained using large-scale unlabeled molecular databases and then fine-tuned for adaption to biomedical text mining. Subsequently tremendous efforts have been devoted to pre-trained language model for biomedical prediction tasks, including property prediction [118], molecular generation [131], peptide and HLA (pHLA) binding prediction [132]. More systematic introduction of molecular pre-trained models refer to [133].

Hybrid methods

Despite the remarkable progress gained by previous methods, improving the prediction accuracy is still crucial. Hybrid methods is proposed to combine with two or multiple types of existing methods in an efficient pattern. For example, GoGNN [22] extracts features from both structured entity graphs and DDI network in a hierarchical way via dual-attention mechanism. MUFFIN [29] jointly learns the drug representation from molecular structure

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

and biomedical KG. BioDKG-DDI [52] adaptively integrates three different types of drug features, molecular structure, drug global information and drug functional similarity representation to predict novel DDIs. Recently, contrastive learning has been successfully applied in the application of bioinformatics, including gene regulatory interactions [134], and drug–target interaction [135]. A novel unsupervised contrastive learning method named MIRACLE [31] is introduced, and it treats a DDI network as a multi-view graph where each node in the interaction graph represents a drug molecular graph instance. AMDE [53] jointly encodes 2D graph feature and 1D SMILES sequence by using message passing attention network and Transformer, respectively.

As illustrated in Table 1 in chronological order, we present different deep and graph learning methods. Specifically, the columns of *Model*, *Input*, *Representation*, *Architecture*, *Task* and *Code* represent the name, data format as the input of model, encoding form of input data, detailed architecture or technology adopted by the proposed model, functions can be implemented by the model, and available link of source code, respectively. The methods listed in Table 1 are also appeared in the taxonomy in Figure 1.

DISCUSSION

To comprehensively investigate the predictive performance of deep and graph learning models, we compared the experimental results of surveyed methods under binary, multi-class and multi-label classification tasks, respectively. In the following sections, we first introduce the benchmark datasets, then present the detailed of evaluation metrics under different prediction tasks, and finally analyze the comparison results.

Benchmark dataset

We chose DrugBank and TWOSIDES datasets as benchmark datasets owing to their wide use across many studies. For the binary classification task, it always assigned a label '1' or '0' to indicate whether an interaction occurs between each pair of drugs in DrugBank and TWOSIDES datasets. For the multi-class classification task, the DrugBank dataset contains 191 808 DDI triplets with 1706 drugs and 86 types of pharmacological relationships between drugs [56]. Following the same criterion in *Decagon* [14], the interaction types with <500 triplets were removed, resulting in 4 576 287 DDI triplets with 963 interaction types in the TWOSIDES dataset. For the multi-label classification task, the TWOSIDES dataset contains 645 drugs (nodes) and 46 221 drug–drug pairs (edges) with 200 different drug side effect types as labels. For each edge, it may be associated with multiple labels. Following *Decagon* [14], it kept 200 commonly occurring DDI types ranging from Top-600 to Top-800 to ensure every DDI type has at least 900 drug combinations. As reported in *KG2ECapsule* [58], it extracted the drug relation from the *description* of *drug-interaction* in DrugBank dataset. The aim was to categorize the types of DDI relations into two groups based on this extracted information. Following the same data split scheme in GMPNN [41], the benchmark dataset was split into train, validation and test sets using a ratio of 6:2:2. Negative samples were randomly generated at a ratio of 1:1, meaning that they consisted of drug pairs that had not appeared in the positive samples.

Evaluation metrics

Generally, we denote the true label and predicted values of DDIs by y and $\hat{y}$, respectively. For the binary classification prediction, experiment results are reported with the following four metrics across the 5-folds. *Area under the precision-recall curve (AUPRC)* is the area under the plot of the precision rate against recall rate at various thresholds, *accuracy (ACC)* is defined as the number of correct predictions divided by the number of total predictions, *area under the receiver operating characteristic (AUROC)* is the area under the plot of the true positive rate against the false positive rate at various thresholds, and *F1 score* is the harmonic mean of precision and recall. The corresponding mathematical calculation is represented as follows.

$$Precision = \frac{TP}{TP + FP}, \tag{21}$$

$$Recall = TPR = \frac{TP}{TP + FN}, \tag{22}$$

$$FPR = \frac{FPR}{FP + TN} \tag{23}$$

$$ACC = \frac{TP + TN}{TN + TP + FN + FP}, \tag{24}$$

$$F1 - score = \frac{2TP}{2TP + FN + FP}, \tag{25}$$

where TP, FP, TN and FN denote the value of true positive, false positive, true negative and false negative, respectively. The *AUPRC* curve is drawn based on the values of FPR and TPR, where the x-axis is TPR and the y-axis is FPR. This is in contrast to *AUROC* curves, where the x-axis is FPR and the y-axis is TPR.

For the multi-class classification prediction, we follow DeepDDI [13] and consider the following metrics, including *mean accuracy*, *macro precision*, *macro recall* and *macro F1*. Macro metrics are used to reflect the average performance across different interaction types. For example, *macro precision* is defined as the average of the precision values of different interaction types. Their definitions are as follows:

$$Mean\ accuracy = \frac{1}{l}\sum_{i=1}^{l}\frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}, \tag{26}$$

$$Macro\ recall = \frac{1}{l}\sum_{i=1}^{l}\frac{TP_i}{TP_i + FN_i}, \tag{27}$$

$$Macro\ precision = \frac{1}{l}\sum_{i=1}^{l}\frac{TP_i}{TP_i + FP_i}, \tag{28}$$

$$Macro\ F1 = \frac{2(Macro\ precision)(Macro\ recall)}{(Macro\ precison) + (Macro\ recall)}, \tag{29}$$

where l is the number of DDI interaction types. In addition, to considering both *precision* and *recall*, we selected the threshold value, which achieves the maximum value of *F1* in each interaction type, as the type-specific threshold.

For multi-label classification prediction, we follow *SumGNN* [30] and a group of metrics is used to measure the prediction, including *ROC-AUC*, *PR-AUC*, *Accuracy (ACC)* and *F1-score*. *ROC–AUC* is the average area under the receiver operating characteristic curve as $ROC - AUC = \sum_{k=1}^{n} TP_k \Delta FP_k$, where k represents kth true-positive and false-positive operating point (TP_k, FP_k). *PR-AUC* is the average area under precision-recall curve $PR - AUC = \sum_{k=1} n \Delta Rec_k$, where k is kth precision/recall oprating point ($Prec_k$, Rec_k). For each side effect type, the performance is individually calculated and

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

use the average performance over all side effects as the final result.

Results

In this section, we compared state-of-the-art deep and graph learning models under binary, multi-class and multi-label classification prediction task, respectively. Table 3 shows the comparison results of 30 models under binary classification task on two benchmark datasets. The performance of DDIs prediction achieved by these models were all measured in terms of AUPR, ACC, AUROC and AUC under 5-fold cross-validation. The greater these evaluation metrics the better the prediction. Although the division of the training and test sets could be specific to models, such an evaluation is still statistically significant. Specifically, from the observation we found that RANEDDI (AUPRC = 0.9894) and KGNN (AUPRC = 0.9892), which belong to network based methods, achieve the best and second-best AUPRC performance compared with chemical structure based and hybrid methods on DrugBank datasets. This is because these methods (i.e. RANEDDI and KGNN) can explore multi-relational information contained in the DDI network or knowledge graph, while the graph embedding approaches like DeepWalk, GraRep, DeepDDI or substructure based method (e.g. CASTER) only learn from similar drug features or chemical structural information. Note that DANN-DDI obtains the best ACC result of 0.9962 over all models, and R^2-DDI achieves the best performance in terms of AUROC and F1 score. Meanwhile, experimental results on TWOSIDES dataset show that DSN-DDI, recently published chemical structure based model, achieves better performance than other baseline models on all evaluation metrics. Particularly, the ACC, AUROC and F1 result of DSN-DDI is 0.9883, 0.9990 and 0.9883, respectively. Interestingly, the comparison indicates that network based methods (e.g. DANN-DDI and RANE-DDI) show similar performance to chemical structure based methods (e.g. R^2-DDI), while hybrid methods show stable performance on DrugBank datasets under binary classification task.

In the multi-class classification task, we chose the DrugBank dataset as benchmark dataset owing to its wide use across many studies and collected 10 deep and graph learning models into the comparison list, which is shown in Table 6. From this table, we found that chemical structure based methods significantly outperform network based and hybrid methods on most metrics. More specifically, Molormer achieved the best score of 0.9667 on mean accuracy, DGNN-DDI achieved the macro recall and macro F1 score of 0.9788 and 0.9616 compared to other methods, respectively. In addition, we can see that chemical structure based methods achieved stable performances across all metrics. For example, the macro recall of GMPNN and SA-DDI are 0.9725 and 0.9746, respectively. These results demonstrate that they achieved similar performance with DGNN-DDI, indicating that they belong to substructure based methods. This is a very encouraging result. The reason could be that (i) DDIs are fundamentally caused by chemical substructure interactions, especially in multi-class classification tasks that focus on atom similarity and key substructures; (ii) more effective strategies are proposed by these methods to specifically detect substructures with irregular size and shape, which can further enhance the representation capability of the model. In addition, the multi-class classification task is more difficult than the binary classification task.

In the multi-label classification task, we chose DrugBank and TWOSIDES datasets as benchmark datasets and compared 11 deep and graph learning models as shown in Table 5. From this table, we observed that KG2ECapsule and SumGNN consistently outperformed other methods in all evaluation metrics. In particular, KG2ECapsule improves over the strongest baselines with respect to PR-AUC by 2.71%, ACC by 1.03%, ROC–AUC by 2.8% and F1 by 2% on DrugBank dataset, respectively. The reason for this is that KG2ECapsule is capable of modeling the triplets and integrating the relations of edges into embedding. Meanwhile, on TWOSIDES dataset, SumGNN achieved at least 2.45% on PR-AUC, 2.82% on ROC-AUC higher performance than other methods. This justifies that SumGNN is more effective to harness the external knowledge via subgraphs. More interestingly, with comparison to other KG based methods (e.g. KGNN and KG-DDI), we found that KG2ECapsule and SumGNN can consistently outperform them on both datasets, which indicates that simply adopting KG embeddings as well as neighborhood sampling are insufficient to fully harness the KG information for DDIs prediction. Moreover, network based methods achieved better performances in the multi-label classification task.

CHALLENGE AND OPPORTUNITIES

Deep and graph learning techniques have distinct advantages over traditional machine learning methods in tackling the computational drug discovery. Although many studies focus on the prediction of DDIs and high prediction performance have been proposed, there still remains several challenges and promising future directions as follows.

Dataset imbalance

Most deep and graph learning models in drug discovery pipeline need large amounts of data for model training and validation. The lack of enough known DDIs and experimentally validated negative samples are major obstacles for deep and graph learning models to have positive influence on DDIs prediction, especially in the application of real-world scenarios. For example, the imbalanced data for different relations in certain case are very sparse with respect to side-effect type, which will lead to poor generalizability for model performance. Meanwhile, current DDI benchmark datasets only include a small number of labeled (resp., positive) samples, in which the quality of data is not guaranteed and the dataset might be imbalanced for the lack of negative samples of drug–drug pair. As for unlabeled samples, most methods regard them as negative samples and sample the same number of negative drug pairs from non-interacting DDIs for model training. These methods overlook the fact that unlabeled samples may contain potential positive data, which would adversely influence the model performances. How to choose high-quality data and how to address insufficient training data remain challenges.

Multimodal representation

The potential of computational drug discovery lies in the variety of multiple data modalities that provide complementary information [136]. Deep and graph learning models using multimodal data will have considerable advantages over unimodal counterparts since the multimodal data offer complementary perspectives. Existing studies usually focus on the single modal data. For example, graph- or substructure-based method pay more attention to the molecular data containing structural information, while network-based approaches only consider the relationship between drug and relation in the drug level, neglecting the atom level of the pair interaction between drugs. These methods do not fully use other data modalities, such as drug–target interactions, drug–disease associations, protein pathways and evidences from electronic medical records, such information may be also highly

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

Table 6. List of widely used platform and toolkit for biomedical application

Name	Year	Release	Updates	Application examples	Website
DeepChem	2017	V2.7.1	√	Molecular property prediction Drug–target binding affinity prediction Physical properties prediction Protein structure analysis and descriptors extraction Number counting of cells in a microscopy image	Link
DeepPurpose	2020	V0.1.5	√	Drug target interaction prediction Drug property prediction Drug–drug interactions prediction Protein-protein interaction prediction Protein function prediction Antiviral drugs repurposing for SARS-CoV2 3CLPro Repurposing using customized training data	Link
PaddleHelix	2020	V1.1.0	√	Large-scale pre-training models of compounds and proteins Molecular property prediction Drug–target affinity prediction Molecular generation RNA design Drug–drug synergy prediction	Link
DGL-LifeSci	2021	V0.3.1	√	Property prediction Generative models Protein-ligand binding affinity prediction Reaction prediction	Link
TorchDrug	2021	V0.2.0	√	Property prediction Pretrained molecular representations Molecule generation Retrosynthesis Knowledge graph reasoning	Link
ADMETlab	2021	V2.0	√	Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) prediction	Link
ChemicalX	2022	V0.1.0	×	Drug–drug interactions prediction Drug pair scoring task	Link

related to DDIs and their induced adverse reactions. Thus, how to effectively utilize diverse and heterogeneous biological data is worth of exploring.

High-order drug associations

Identifying the potential associations between drugs and related entities (e.g. diseases and microbe) is pivotal to understanding the underlying disease mechanisms and facilitating personalized treatments. In the past few years, most methods have been proposed to concentrate on predicting pair-wise associations, such as drug–drug, drug–protein, drug–microbe and drug–disease interactions, these methods deal with them separately and fail to provide in-depth insights into high-order association patterns. For example, many diseases are closely related to various microbes, which interact with a variety of drugs in complex way, and the causal links between drugs, gut microbes and diseases require a workflow to uncover their intricate interactions. Such a workflow of triple-wise drug-microbe-disease associations can be regarded as high-order drug associations prediction. Meanwhile, high-order associations prediction is a fundamental task in multiple domains, including knowledge graphs, recommendation systems and bioinformatics. There is an urgent need to seek ways to develop effective methods for predicting high-order drug associations to speed up the process of drug discovery.

Model interpretability

Deep and graph learning techniques offer great potential in many fields, but they are often essentially 'black boxes' that are unable to provide confidence and actionability for the predicted results. As an essential process in drug discovery, DDIs prediction aims to identify and quantify the risks related to the usage of drugs for a better understanding of adverse drug effects and the pathogenic mechanisms. The latent embedding obtained by current deep and graph learning models is limited to capturing implicit correlations of the data, which is hard to provide reasonable explanations for the predicted interactions. Thus, the idea model should understand how the algorithms are constructed, what each layer learns, and what the embeddings represent. Meanwhile, interpretability and evidence support are essential for prediction methods in biomedical applications. It is also worthwhile to further focus on interpretability and to improve the reliability of predicted results.

Generative AI models

Recent advances in generative AI models, such as ChatGPT (https://openai.com/blog/chatgpt), have shown remarkable success on a variety of domains. From Transformer to BERT to ChatGPT, the continuous advancement of generative AI models has opened up a new era of AI. These generative AI models are trained on large-scale datasets, providing a reasonable parameter initialization for a wide range of downstream applications, including natural language processing [137], computer vision [138] and graph learning [139]. Moreover, generative AI models have been deployed in various stages of the drug development pipeline [10], ranging from AI-assisted target selection and validation to molecular design and chemical synthesis. In the near future, it is anticipated that generative AI models would

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

be able to generate realistic data that can be used to identify potential DDIs. These data can then be utilized to improve existing models or create new models that are more effective at addressing the challenges mentioned above. By combining the power of generative AI models and advanced deep and graph learning techniques, it is conceivable to develop better models for predicting DDIs.

Platform and toolkit

To further speed up the drug discovery process and enable more people with different scientific backgrounds to get involved in research, many researcher and communities have been committed to the development of platform and toolkit based on machine learning and deep learning methods. Table 6 illustrates the widely used platform and toolkit for biomedical application. Specifically, DeepChem aims to provide a high quality open-source toolchain that makes deep learning in drug discovery, materials science, quantum chemistry, and biology more accessible. DGL-LifeSci is a DGL-based package for various life science applications with graph neural networks and provides various functions. DeepPurpose is a deep learning-based molecular modeling and prediction toolkit involving many downstream tasks (e.g. compound property prediction and protein function prediction). TorchDrug is a machine learning platform designed for drug discovery that covers various techniques from GNNs, geometric deep learning, KGs, deep generative models, and reinforcement learning. PaddleHelix is a bio-computing tool that takes advantage of the machine learning approaches, especially DNNs, for facilitating the development of the following areas, including drug discovery, vaccine design and precision medicine. ADMETlab 2.0 [140] is an improved version of the widely used ADMETlab, which is used to systematical evaluation of ADMET properties. While fewer work are specialized for developing the platform or toolkit on DDIs prediction. To our knowledge, one such work named ChemicalX is a deep learning library for drug–drug interaction, polypharmacy side effect and synergy prediction, and also includes state-of-the-art DNN architectures that solve the drug pair scoring task, with implemented methods covering traditional SMILES sequence based techniques and MPNN based models. However, these platforms and toolkits, which are mainly developed by individuals, do not have any maintenance or update schedule in place. As a result, they will become increasingly obsolete as the underlying programming framework and deep learning models continue to evolve.

CONCLUSIONS AND OUTLOOK

In this work, we provided a comprehensive review of deep and graph learning methods for drug–drug interactions prediction. We categorized existing approaches into traditional machine learning, deep learning and GNN-based methods. We introduced data sources and summarized the widely used molecular representation as well as some classic GNN model on DDIs prediction. To the end, we discussed the current challenges of existing deep and graph learning methods and suggested potential research directions for further development in DDIs prediction. In conclusion, the rapidly growth of deep and graph learning techniques has brought new opportunities for biomedical applications, including drug-related prediction tasks. However, the bottlenecks of these technologies, such as imbalance dataset, the issues of multimodal representation and high-order drug associations prediction, and the lack of or limited interpretability of the model impedes their application and further affects their prediction performance. Therefore, there is an urgent need to further develop and evaluate intelligent deep and graph learning models in realistic drug discovery scenarios in order to reach its full potential.

Key Points

- *Structured taxonomy*. As shown in Figure 1, we contribute a structured taxonomy to provide a broad overview of computational methods, which categorizes existing works from four perspectives: chemical structure based, network based, NLP based and hybrid methods.
- *Current progress*. We systematically delineate the current research directions on the topic of deep and graph learning methods for DDIs prediction as illustrated in Table 1, and we further investigate the comparison performance of these representative baseline models as shown in Tables 3–5.
- *Abundant resources*. We have gathered a comprehensive collection of resources dedicated to DDIs prediction. These collections include open-sourced deep and graph learning methods, available platform and toolkit, as well as an important paper list. These resources can be accessed our github (https://github.com/xzenglab/resources-for-DDIs-prediction-using-DL), which will be continuously updated.
- *Future directions*. We discuss the limitations of existing works and suggest several promising future directions.

FUNDING

We thank the editors and reviewers for their efforts in reviewing this manuscript. The work is supported in part by the National Natural Science Foundation of China (Nos 62202413, 62272151, 61972138, U21A20427), the National Key Researchand Development Program of China (No. 2020YFC0832405), the Science and Technology Innovation Program of Hunan Province of China (No. 2022RC1099, 2022WK2009), the High-Level Talent Aggregation Project in Hunan Province, China Innovation Team (No. 2019RS1060), the Hunan Provincial Natural Science Foundation of China under Grants (No. 2022JJ20016, 2022JJ40451, 2023JJ30161), the Science and Technology Innovation 2030-Major Project (No. 2021ZD0150100), the General Project of Hunan Provincial Education Department (No. 21C0074, 22A0022), the Open Research Projects of Zhejiang Lab (2021RD0AB02), the Natural Science Foundationof Changsha City (No. kq2202137), and the NSF under Grants III-1763325, III-1909323, III-2106758 and SaTC-1930941.

REFERENCES

1. Kantor ED, Rehm CD, Mengmeng D, *et al.* Trends in dietary supplement use among us adults from 1999-2012. *JAMA* 2016;**316**(14):1464–74.
2. Jin B, Yang H, Xiao C, *et al.* Multitask dyadic prediction and its application in prediction of adverse drug–drug interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI press, Vol. **31**, 2017.
3. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *J Am Med Inform Assoc* 2012;**19**(1):79–85.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

4. Vilar S, Uriarte E, Santana L, *et al.* Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc* 2014;**9**(9):2147–63.
5. Bansal M, Yang J, Karan C, *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 2014;**32**(12):1213–22.
6. Safdari R, Ferdousi R, Aziziheris K, *et al.* Computerized techniques pave the way for drug-drug interaction prediction and interpretation. *Bioimpacts Bi* 2016;**6**(2):71–8.
7. Qiu Y, Zhang Y, Deng Y, *et al.* A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM Trans Comput Biol Bioinform* 2021.
8. Dong J, Zhao M, Liu Y, *et al.* Deep learning in retrosynthesis planning: datasets, models and tools. *Brief Bioinform* 2022;**23**(1):bbab391.
9. Pan X, Lin X, Cao D, *et al.* Deep learning for drug repurposing: methods, databases, and applications. *Wiley Interdiscipl Rev: Comput Mol Sci* 2022;**12**(4):e1597.
10. Zeng X, Wang F, Luo Y, *et al.* Deep generative molecular design reshapes drug discovery. *Cell Rep Med*. Cell press, 2022, 100794.
11. Zeng X, Xinqi T, Liu Y, *et al.* Toward better drug discovery with knowledge graph. *Curr Opin Struct Biol* 2022;**72**:114–26.
12. Zhang T, Leng J, Liu Y. Deep learning for drug–drug interaction extraction from the literature: a review. *Brief Bioinform*. Oxford, 2020;**21**(5):1609–27.
13. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc Natl Acad Sci* 2018;**115**(18):E4304–11.
14. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**(13):i457–66.
15. Xu C, Lin Y, Wang Y, *et al.* Mlrda: A multi-task semi-supervised learning framework for drug-drug interaction prediction. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 4518–24, 2019.
16. Shen Y, Yuan K, Yang M, *et al.* Kmr: knowledge-oriented medicine representation learning for drug–drug interaction and similarity computation. *J Chem*. Springer, 2019;**11**(1):1–16.
17. Xu N, Wang P, Chen L, *et al.* Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 3968–74. International Joint Conferences on Artificial Intelligence Organization, **7**, 2019.
18. Deac A, Huang Y-H, Veličković P, *et al.* Drug-drug adverse effect prediction with graph co-attention. arXiv preprint arXiv:1905.00534. 2019.
19. Peng B, Ning X. Deep learning for high-order drug-drug interaction prediction. In: *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 197–206, 2019.
20. Karim MR, Cochez M, Jares JB, *et al.* 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*, 113–23, 2019.
21. Deng Y, Xinran X, Qiu Y, *et al.* A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 2020;**36**(15):4316–22.
22. Wang H, Lian D, Zhang Y, Lu Qin, *et al.* Gognn: Graph of graphs neural network for predicting structured entity interactions. In: Bessiere C (ed). *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, 1317–23. International Joint Conferences on Artificial Intelligence Organization, **7**, 2020. Main track.
23. Lin X, Quan Z, Wang Z-J, *et al.* Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In: *International Joint Conferences on Artificial Intelligence*. 2020;**380**:2739–45.
24. Lee J, Yoon W, Kim S, *et al.* Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;**36**(4):1234–40.
25. Huang K, Xiao C, Glass LM, *et al.* Skipgnn: predicting molecular interactions with skip-graph networks. *Sci Rep* 2020;**10**(1):1–16.
26. Huang K, Xiao C, Hoang T, *et al.* Caster: predicting drug interactions with chemical substructure representation. *In: Proceedings of the AAAI Conference on Artificial Intelligence* 2020; 702–9.
27. Sun M, Wang F, Elemento O, *et al.* Structure-based drug-drug interaction detection via expressive graph convolutional networks and deep sets (student abstract). In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 13927–8, 2020.
28. Kishan KC, Li R, Cui F, Haake AR. Predicting biomedical interactions with higher-order graph convolutional networks. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**(2):676–87.
29. Chen Y, Ma T, Yang X, *et al.* Muffin: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics*. Oxford, 2021;**37**(17):2651–8.
30. Yue Y, Huang K, Zhang C, *et al.* Sumgnn: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* 2021;**37**(18):2988–95.
31. Wang Y, Min Y, Chen X, Ji W. Multi-view graph contrastive representation learning for drug-drug interaction prediction. *In: Proceedings of the Web Conference* 2021;**2021**:2921–33.
32. Nyamabo AK, Hui Y, Shi J-Y. Ssi-ddi: substructure-substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;**22**(6):bbab133.
33. Dai Y, Guo C, Guo W, Eickhoff C. Drug–drug interaction prediction with Wasserstein adversarial autoencoder-based knowledge graph embeddings. *Brief Bioinform* 2021;**22**(4):bbaa256.
34. Feng Y-H, Zhang S-W. Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs. *Molecules*. MDPI, 2022;**27**(9):3004.
35. He C, Liu Y, Li H, *et al.* Multi-type feature fusion based on graph neural network for drug-drug interaction prediction. *BMC Bioinformatics* 2022;**23**(1):224.
36. Wang F, Lei X, Liao B, Fang-Xiang W. Predicting drug–drug interactions by graph convolutional network with multi-kernel. *Brief Bioinform* 2022;**23**(1):bbab511.
37. Yin Q, Cao X, Fan R, *et al.* Deepdrug: a general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. biorxiv. 2022;2020–11.
38. Kang C, Zhang H, Liu Z, *et al.* Lr-gnn: a graph neural network based on link representation for predicting molecular associations. *Brief Bioinform* 2022;**23**(1):bbab513.
39. Liu S, Zhang Y, Cui Y, *et al.* Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**20**(0):976–85.
40. Feng Y-Y, Yu H, Feng Y-H, *et al.* Directed graph attention networks for predicting asymmetric drug–drug interactions. *Brief Bioinform* 2022;**23**(3):bbac151.
41. Nyamabo AK, Hui Y, Liu Z, Shi J-Y. Drug-drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinform* 2022;**23**(1):bbab441.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

42. Hui Y, Zhao SY, Shi JY. Stnn-ddi: a substructure-aware tensor neural network to predict drug–drug interactions. *Brief Bioinform* 2022;**23**(4):bbac209.
43. Feng Y-H, Zhang S-W, Zhang Q-Q, *et al.* Deepmddi: a deep graph convolutional network framework for multi-label prediction of drug-drug interactions. *Anal Biochem* 2022;**646**:114631.
44. Hui Y, Dong WM, Shi JY. Raneddi: relation-aware network embedding for drug-drug interaction prediction. *Inform Sci* 2022;**582**:167–80.
45. Kim E, Nam H. Deside-ddi: interpretable prediction of drug-drug interactions using drug-induced gene expressions. *J Chem* 2022;**14**(1):1–12.
46. Yang Z, Zhong W, Lv Q, Chen CY-C. Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network. *Chem Sci* 2022;**13**(29):8693–703.
47. Zhu X, Shen Y, Weiming L. Molecular substructure-aware network for drug-drug interaction prediction. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, 4757–61.
48. Hong Y, Luo P, Jin S, Liu X. Lagat: link-aware graph attention network for drug-drug interaction prediction. *Bioinformatics* 2022.
49. Zhang X, Wang G, Meng X, *et al.* Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug–drug interactions prediction. *Brief Bioinform* 2022;**23**(5):bbac296.
50. Lin S, Chen W, Chen G, *et al.* Mddi-scl: predicting multi-type drug-drug interactions via supervised contrastive learning. *J Chem* 2022;**14**(1):1–12.
51. Lin J, Lijun W, Zhu J, *et al.* R2-ddi: relation-aware feature refinement for drug–drug interaction prediction. *Brief Bioinform* 2023;**24**(1):bbac576.
52. Ren Z-H, Chang-Qing Y, Li L-P, *et al.* Biodkg–ddi: predicting drug–drug interactions based on drug knowledge graph fusing biochemical information. *Brief Funct Genomics* 2022;**21**(3):216–29.
53. Pang S, Zhang Y, Song T, *et al.* Amde: a novel attention-mechanism-based multidimensional feature encoder for drug–drug interaction prediction. *Brief Bioinform*. Oxford, 2022;**23**(1):bbab545.
54. Xiaorui S, Lun H, You Z, *et al.* Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Brief Bioinform* 2022;**23**(3):bbac140.
55. He H, Chen G, Chen CY-C. 3dgt-ddi: 3d graph and text based neural network for drug–drug interaction prediction. *Brief Bioinform*. Oxford, 2022;**23**(3):bbac134.
56. Li Z, Zhu S, Shao B, *et al.* Dsn-ddi: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Brief Bioinform* 2023.
57. Ma M, Lei X. A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLoS Comput Biol* 2023;**19**(1):e1010812.
58. Xiaorui S, You Z-H, Huang D-s, *et al.* Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction. *IEEE Trans Knowl Data Eng* 2022.
59. Kanehisa M, Furumichi M, Sato Y, *et al.* Kegg for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 2023;**51**(D1):D587–92.
60. Wishart DS, Feunang YD, Guo AC, *et al.* Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res*. Oxford, 2018;**46**(D1):D1074–82.
61. Kuhn M, Letunic I, Jensen LJ, Bork P. The sider database of drugs and side effects. *Nucleic Acids Res* 2016;**44**(D1):D1075–9.
62. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**(125):125ra31.
63. Zitnik M, Sosic R, Leskovec J. Biosnap datasets: Stanford biomedical network dataset collection. Note: http://snap. stanford. edu/biodata Cited by, **5**(1), 2018.
64. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, 701–10.
65. Cao S, Lu W, Qiongkai X. Grarep: Learning graph representations with global structural information. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, –900, 2015.
66. Tang J, Meng Qu, Wang M, Zhang M, *et al.* Line: Large-scale information network embedding. In: *Proceedings of the 24th International Conference on World Wide Web*, 1067–77, 2015.
67. Wang D, Cui P, Zhu W. Structural deep network embedding. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016;**1**:1225–34.
68. Kipf TN, Welling M. Variational graph auto-encoders. arXiv preprint. arXiv:1611.07308. 2016.
69. LFR Ribeiro, Saverese PHP, Figueiredo DR. struc2vec: Learning node representations from structural identity. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 385–94, 2017.
70. Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiol Drug Saf* 2004;**13**(8):519–23.
71. Baldessarini RJ. In: Brunton LL, Lazo JS, Parker KL (eds). *Drug Therapy of Depression and Anxiety Disorders. Goodman and Gilmans The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill, 2006, 429–60.
72. Weininger D. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**(1):31–6.
73. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. *Adv Neural Inform Process Syst* 2014;**27**:1–9.
74. Xu Z, Wang S, Zhu F, *et al.* Seq2seq fingerprint: an unsupervised deep molecular embedding for drug discovery. In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 285–94, 2017.
75. Zhang X, Wang S, Zhu F, *et al.* Seq3seq fingerprint: towards end-to-end semi-supervised deep drug discovery. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 404–13, 2018.
76. Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J Chem Inf Model* 2018;**58**(1):27–35.
77. Huang K, Xiao C, Glass LM, Sun J. Moltrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics* 2021;**37**(6):830–6.
78. Landrum G, *et al.* Rdkit: open-source cheminformatics. 2006.
79. Gilmer J, Schoenholz SS, Riley PF, *et al.* Neural message passing for quantum chemistry. In: *International Conference on Machine Learning*. PMLR, 2017, 1263–72.
80. Li S, Zhou J, Tong Xu, Dou D, Xiong H. Geomgcl: geometric graph contrastive learning for molecular property prediction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. **36**, 4541–9, 2022.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

81. Ganea O, Pattanaik L, Coley C, *et al.* Geomol: torsional geometric generation of molecular 3d conformer ensembles. *Adv Neural Inform Process Syst* 2021;**34**:13757–69.
82. Zhang C, Song D, Huang C, *et al.* Heterogeneous graph neural network. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, 793–803.
83. Ji S, Pan S, Cambria E, *et al.* A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans Neural Netw Learn Syst* 2021;**33**(2):494–514.
84. Domingo-Fernández D, Baksi S, Schultz B, *et al.* Covid-19 knowledge graph: a computable, multi-modal, cause-and-effect knowledge model of covid-19 pathophysiology. *Bioinformatics* 2021;**37**(9):1332–4.
85. Reese JT, Unni D, Callahan TJ, *et al.* Kg-covid-19: a framework to produce customized knowledge graphs for covid-19 response. *Patterns* 2021;**2**(1):100155.
86. Himmelstein DS, Lizee A, Hessler C, *et al.* Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife* 2017;**6**:e26726.
87. Ioannidis VN, Song X, Manchanda S, *et al. Drkg - drug repurposing knowledge graph for covid-19*. https://github.com/gnn4dr/DRKG/, 2020.
88. Walsh B, Mohamed SK, Nováček V. Biokg: A knowledge graph for relational learning on biological data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020, 3173–80.
89. Zheng S, Rao J, Song Y, *et al.* Pharmkg: a dedicated knowledge graph benchmark for biomedical data mining. *Brief Bioinform* 2021;**22**(4):bbaa344.
90. Breit A, Ott S, Agibetov A, Samwald M. Openbiolink: a benchmarking framework for large-scale biomedical link prediction. *Bioinformatics*. Oxford, 2020;**36**(13):4097–8.
91. Santos A, Colaço AR, Nielsen AB, *et al.* Clinical knowledge graph integrates proteomics data into clinical decision-making. bioRxiv. 2020.
92. Bonner S, Barrett IP, Ye C, *et al.* A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Brief Bioinform* 2022;**23**(6):bbac404.
93. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. 2016.
94. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. *Adv Neural Inform Process Syst* 2017;**30**.
95. Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. In: *International Conference on Learning Representations (ICLR)*, 2018.
96. Gainza P, Sverrisson F, Monti F, *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods*. Springer Nature, 2020;**17**(2):184–92.
97. Méndez-Lucio O, Baillif B, Clevert D-A, *et al.* De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat Commun* 2020;**11**(1):1–10.
98. Beker W, Wołos A, Szymkuć S, Grzybowski BA. Minimal-uncertainty prediction of general drug-likeness based on Bayesian neural networks. *Nat Mach Intell* 2020;**2**(8):457–65.
99. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with alphafold. *Nature* 2021;**596**(7873):583–9.
100. Zeng X, Zhu S, Liu X, *et al.* Deepdr: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 2019;**35**(24):5191–8.
101. Zeng X, Zhu S, Weiqiang L, *et al.* Target identification among known drugs by deep learning from heterogeneous networks. *Chem Sci* 2020;**11**(7):1775–97.
102. Yang X, Niu Z, Liu Y, *et al.* Modality-DTA: multimodality fusion strategy for drug-target affinity prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**20**(2):1200–10.
103. Zeng X, Xiang H, Linhui Y, *et al.* Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat Mach Intell* 2022;1–13.
104. Ma T, Lin X, Song B, *et al.* Kg-mtl: knowledge graph enhanced multi-task learning for molecular interaction. *IEEE Trans Knowl Data Eng* 2022.
105. Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics* 2019;**20**(1):1–8.
106. Lin S, Wang Y, Zhang L, *et al.* Mdf-sa-ddi: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief Bioinform* 2022;**23**(1):bbab421.
107. Kim Y, Meystre SM. Ensemble method–based extraction of medication and related information from clinical texts. *J Am Med Inform Assoc* 2020;**27**(1):31–8.
108. Datta A, Flynn NR, Barnette DA, *et al.* Machine learning liver-injuring drug interactions with non-steroidal anti-inflammatory drugs (nsaids) from a retrospective electronic health record (ehr) cohort. *PLoS Comput Biol* 2021;**17**(7):e1009053.
109. Patrick W, Nelson SD, Zhao J, *et al.* Ddiwas: high-throughput electronic health record-based screening of drug-drug interactions. *J Am Med Inform Assoc* 2021;**28**(7):1421–30.
110. Vilar S, Friedman C, Hripcsak G. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform* 2018;**19**(5):863–77.
111. Vilar S, Harpaz R, Uriarte E, *et al.* Drug-drug interaction through molecular structure similarity analysis. *J Am Med Inform Assoc* 2012;**19**(6):1066–74.
112. Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J Am Med Inform Assoc* 2014;**21**(e2):e278–86.
113. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform* 2017;**70**:54–64.
114. Huang H, Zhang P, Qu XA, *et al.* Systematic prediction of drug combinations based on clinical side-effects. *Sci Rep* 2014;**4**(1):1–7.
115. Wang Y, Wang J, Cao Z, Farimani AB. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell* 2022;**4**(3):279–87.
116. Li Y, Hsieh C-Y, Ruiqiang L, *et al.* An adaptive graph learning method for automated molecular interactions and properties predictions. *Nat Mach Intell* 2022;**4**(7):645–51.
117. Duvenaud DK, Maclaurin D, Iparraguirre J, *et al.* Convolutional networks on graphs for learning molecular fingerprints. *Adv Neural Inform Process Syst* 2015;**28**.
118. Fang X, Liu L, Lei J, *et al.* Geometry-enhanced molecular representation learning for property prediction. *Nat Mach Intell* 2022;**4**(2):127–34.
119. Zhang P, Wang F, Jianying H, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**(1):1–10.
120. Park K, Kim D, Ha S, Lee D. Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks. *PloS One* 2015;**10**(10):e0140816.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025

121. Cami A, Manzi S, Arnold A, Reis BY. Pharmacointeraction network models predict unknown drug-drug interactions. *PloS One* 2013;**8**(4):e61468.
122. Zhang W, Jing K, Huang F, *et al.* Sflln: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inform Sci* 2019;**497**:189–201.
123. Mingdong Ou, Cui P, Pei J, Zhang Z, *et al.* Asymmetric transitivity preserving graph embedding. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1105–14, 2016.
124. Zhang W, Liu X, Chen Y, *et al.* Feature-derived graph regularized matrix factorization for predicting drug side effects. *Neurocomputing* 2018;**287**:154–62.
125. Shi J-Y, Mao K-T, Hui Y, Yiu S-M. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *J Chem* 2019;**11**(1):1–16.
126. Grover A and Leskovec J. node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–64, 2016.
127. Yue X, Wang Z, Huang J, *et al.* Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics* 2020;**36**(4):1241–51.
128. Lun H, Zhang J, Pan X, *et al.* Hiscf: leveraging higher-order structures for clustering analysis in biological networks. *Bioinformatics* 2021;**37**(4):542–50.
129. Wang H, Zhao M, Xie X, *et al.* Knowledge graph convolutional networks for recommender systems. In: *The World Wide Web Conference* 2019, 3307–13.
130. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint. arXiv:1810.04805. 2018.
131. Bagal V, Rishal Aggarwal PK, Vinod, and U Deva Priyakumar. Molgpt: molecular generation using a transformer-decoder model. *J Chem Inf Model* 2021;**62**(9):2064–76.
132. Chu Y, Zhang Y, Wang Q, *et al.* A transformer-based model to predict peptide–hla class i binding and optimize mutated peptides for vaccine design. *Nat Mach Intell* 2022;**4**(3):300–11.
133. Xia J, Zhu Y, Yuanqi D, *et al.* A systematic survey of molecular pre-trained models. arXiv preprint. arXiv:2210.16484. 2022.
134. Zheng L, Liu Z, Yang Y, Shen H-B. Accurate inference of gene regulatory interactions from spatial gene expression with deep contrastive learning. *Bioinformatics* 2022;**38**(3):746–53.
135. Li Y, Qiao G, Gao X, Wang G. Supervised graph co-contrastive learning for drug–target interaction prediction. *Bioinformatics* 2022;**38**(10):2847–54.
136. Luo Y, Eran A, Palmer N, *et al.* A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nat Med* 2020;**26**(9):1375–9.
137. Brown T, Mann B, Ryder N, *et al.* Language models are few-shot learners. *Adv Neural Inform Process Syst* 2020;**33**:1877–901.
138. Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint. arXiv:2010.11929. 2020.
139. Yun S, Jeong M, Kim R, *et al.* Graph transformer networks. *Adv Neural Inform Process Syst* 2019;**32**.
140. Xiong G, Zhenxing W, Yi J, *et al.* Admetlab 2.0: an integrated online platform for accurate and comprehensive predictions of admet properties. *Nucleic Acids Res* 2021;**49**(W1):W5–14.

Downloaded from https://academic.oup.com/bib/article/24/4/bbad235/7217116 by National Science & Technology Library user on 07 March 2025