OXFORD

HTCL-DDI: a hierarchical triple-view contrastive learning framework for drug–drug interaction prediction

Ran Zhang, Xuezhi Wang, Pengfei Wang, Zhen Meng, Wenjuan Cui and Yuanchun Zhou

Corresponding author. Pengfei Wang, Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100083, China. E-mail: pfwang@cnic.cn

Abstract

Drug–drug interaction (DDI) prediction can discover potential risks of drug combinations in advance by detecting drug pairs that are likely to interact with each other, sparking an increasing demand for computational methods of DDI prediction. However, existing computational DDI methods mostly rely on the single-view paradigm, failing to handle the complex features and intricate patterns of DDIs due to the limited expressiveness of the single view. To this end, we propose a <u>H</u>ierarchical <u>T</u>riple-view <u>C</u>ontrastive <u>L</u>earning framework for <u>D</u>rug–<u>D</u>rug Interaction prediction (HTCL-DDI), leveraging the molecular, structural and semantic views to model the complicated information involved in DDI prediction. To aggregate the intra-molecular compositional and structural information, we present a dual attention-aware network in the **molecular view**. Based on the molecular view, to further capture inter-molecular information, we utilize the one-hop neighboring information and high-order semantic relations in the **structural view** and **semantic view**, respectively. Then, we introduce contrastive learning to enhance drug representation learning from multifaceted aspects and improve the robustness of HTCL-DDI. Finally, we conduct extensive experiments on three real-world datasets. All the experimental results show the significant improvement of HTCL-DDI over the state-of-the-art methods, which also demonstrates that HTCL-DDI opens new avenues for ensuring medication safety and identifying synergistic drug combinations.

Keywords: drug–drug interaction prediction, multi-view, graph contrastive learning, graph attention network

INTRODUCTION

Drug–drug interaction (DDI) exerts a significant impact on treatment effectiveness and poses a potential risk to patient safety in the multi-drug combination strategies. Multiple drugs interact with each other and can substantially alter their pharmacological effects, resulting in efficacy changes and even side effects. According to the researches [1, 2], elderly patients in the United States typically take an average of 10 medications, and over half of elderly individuals aged 80 and above in Europe use more than six medications. In China, the number of drugs taken by patients over 80 years old ranges from 8 to 60 [3]. With an aging global population and a corresponding increase in polypharmacy, the incidence of harmful and potentially fatal side effects of DDIs is expected to rise. Therefore, DDI prediction is becoming increasingly important in providing multi-drug combination strategies and avoiding adverse reactions for systematic treatment. However, the vast body of possible DDIs makes it labor-intensive and time-consuming to screen DDIs through wet-lab experiments [4], which provokes an urgent demand for computational DDI prediction techniques.

In general, current computational methods to identify DDIs can be classified into three categories: literature-based methods, similarity-based methods and network-based methods [5].

Literature-based methods extract DDIs from unstructured text sources such as drug instructions, electronic medical records and biomedical literature [6, 7]. Zhang *et al.* [8] integrated side effects extracted from package inserts of prescription drugs and FDA Adverse Event Reporting System to recognize DDIs via a label propagation framework [8]. Despite the wealth of DDIs documented in the vast literature, literature-based methods are limited by the diversity of linguistic expressions and cannot detect newly discovered interactions.

Similarity-based methods operate on the assumption that if drug i interacts with drug j, then drugs that are similar to drug i are also likely to interact with drug j. The key to similarity-based methods lies in how to measure the similarity between drugs [9]. Vilar *et al.* [10] represented the molecular structure as a bit vector to encode the presence or absence of molecular features. Ferdousi *et al.* [11] utilized the Rus–Rao approach with 12 binary vectors to calculate the similarity of drug pairs. Matrix

Ran Zhang is a PhD student in Computer Network Information Center, Chinese Academy of Sciences and University of Chinese Academy of Sciences. Her expertise is computational biology and artificial intelligence.
Xuezhi Wang is a researcher in Computer Network Information Center, Chinese Academy of Sciences and University of Chinese Academy of Sciences. His expertise is biological data management and analysis.
Pengfei Wang is an associate researcher in Computer Network Information Center, Chinese Academy of Sciences and University of Chinese Academy of Sciences. His expertise is data mining and computational biology.
Zhen Meng is a senior engineer in Computer Network Information Center, Chinese Academy of Sciences. Her expertise is biological data management and analysis.
Wenjuan Cui is an associate researcher in Computer Network Information Center, Chinese Academy of Sciences and University of Chinese Academy of Sciences. Her expertise is data mining and bioinformatics.
Yuanchun Zhou is a researcher in Computer Network Information Center, Chinese Academy of Sciences and University of Chinese Academy of Sciences. His expertise is data mining and computational biology.

factorization methods can decompose the DDI matrix into several latent feature matrices to discover the drug similarity information [12–15]. Yu *et al*. [16] proposed a DDI-non-negative matrix factorization approach to predict conventional and synthetic DDIs based on semi-non-negative matrix factors. Obviously, similarity-based methods heavily rely on visible data and generally have poor generalization performance.

Network-based methods construct undirected graphs by utilizing the drug structures or known DDIs [17, 18] as molecular structures and DDI associations have natural network structures. For example, Yan *et al*. [19] adopted the node-based drug network diffusion method and recursive least squares algorithm to deduce new DDIs. With the emergence of deep learning methods, graph representation learning has demonstrated groundbreaking performance and great potential for the DDI prediction task [20–27]. Graph embedding methods transform a network into a low-dimensional space while preserving the structural and semantic information [28]. Marinka *et al*. [29] proposed a graph convolutional network (GCN) architecture to identify the type of DDIs. Chen *et al*. [30] introduced a graph neural network (GNN) into DDI prediction and recognized the most vital local atoms via a self-contained attention mechanism [30]. Nyamabo *et al*. [31] used a multi-layer graph attention network (GAT) to extract different substructures of drugs and aggregated them to obtain molecule-level representations of drugs. He *et al*. [32] employed the 3D molecular graph structure and positional information to comprehensively investigate the impact of drug substructure on DDIs.

Drug molecules and DDIs exhibit typical graph structures, and network-based methods introducing graph representation learning have shown outstanding performance advantages. However, it seems probable that existing methods have not fully exploited the rich information inherent in drug molecules and DDIs yet. How to discover and balance the potential relation between drug molecular information and DDI information remains a current research focus. Most network-based methods adhere to the single-view paradigm, learn the single drug representation and optimize the entire model with a supervised loss. The insufficient expressiveness of models is partly attributed to the failure to capture the multi-level information. Inspired by MIRACLE [33], it is significant to integrate multiple views to improve DDI prediction performance, and researchers have also explored multi-view representation learning in the field of DDI prediction [34–39]. Even so, existing multi-view methods generally obtain drug representations from multiple perspectives and perform feature fusion for DDI prediction, without considering intra-molecular and inter-molecular information simultaneously or delving into the mutual enhancement among multi-view representations. In addition, the detailed molecular compositional information and the rich semantic information in the DDI network are often prone to be disregarded.

To tackle the above issues, we propose a novel DDI prediction framework, **H**ierarchical **T**riple-view **C**ontrastive **L**earning framework for **D**rug–**D**rug **I**nteraction prediction (HTCL-DDI). HTCL-DDI unleashes the power of multi-view contrastive learning framework to generate informative drug representations for DDI prediction. Specifically, we construct three views in a two-tier architecture, including a molecular view for drug molecular structure, a structural view for one-hop DDIs and a semantic view for high-order relations. Firstly, the molecular view employs a dual attention-aware message passing network (DAMPN) to embed drug molecular graphs into drug low-dimensional vectors. Subsequently, the structural view applies a single-layer multi-head GAT to aggregate one-hop neighboring structural information,

and the semantic view adopts a three-layer multi-head GAT with topological information (TopoGAT) to capture topological details and semantic information for DDI prediction. Finally, a contrastive learning approach is employed to enhance and balance information from different views for accurate DDI prediction.

Overall, our major contributions of this work can be summarized as follows:

- We construct a two-tier framework with triple views to describe complicated DDI patterns and capture hierarchical information from different perspectives including the molecular, structural and semantic views. HTCL-DDI uses contrastive learning to facilitate mutual enhancement across three views and improve the capability to handle both local and global information.
- We design a novel dual-attention mechanism in the molecular view to generate informative molecular representations and introduce topological information into GAT in the semantic view to model high-order relations, fully exploiting topological correlation to mitigate the intrinsic over-smoothing problem in GAT.
- Extensive experiments conducted on three real-world datasets demonstrate that HTCL-DDI achieves significant performance improvements over the state-of-the-art DDI prediction methods.

METHODOLOGY

As shown in Figure 1, we propose an end-to-end DDI prediction framework with three hierarchical views to preserve more fine-grained and comprehensive DDI information. At the intra-molecular level, we learn drug representations incorporating the atoms and bonds within drug molecules. At the inter-molecular level, we subsequently employ a single-layer multi-head GAT and a three-layer multi-head TopoGAT to capture the structures and semantics of the DDI network, respectively. Finally, we devise a DDI predictor to identify missing links in the DDI network, utilizing contrastive learning to facilitate the proposed model to differentiate between positive and negative drug pairs.

Molecular view

The drug molecular graph determines pharmacological properties and largely affects how the drug interacts with targets in the body. However, it is challenging to directly utilize the molecular graph for DDI prediction, so we embed the drug molecular graph into a low-dimensional vector space. We utilize simplified molecular input line entry system (SMILES) [40] representations as descriptors of drug molecules. The molecular properties are highly dependent on the molecular structure and composition. Inspired by message passing neural networks (MPNNs) [41], we develop a dual attention-aware message passing network (DAMPN) to encode drug molecules. Specifically, we perform message passing between atoms through chemical bonds and aggregate the features of all atoms and bonds within the drug molecule.

A drug molecule can be represented as a graph where each atom is regarded as a node and each bond as an edge. We convert SMILES representations to molecular graphs via RDKit [42]. Thereafter, we extract the atoms and construct the multi-channel adjacency matrix depicting different types of chemical bonds.

For each chemical bond, DAMPN employs a mapping function that projects bond features onto a hidden layer and subsequently applies a nonlinear transformation to produce a message. The
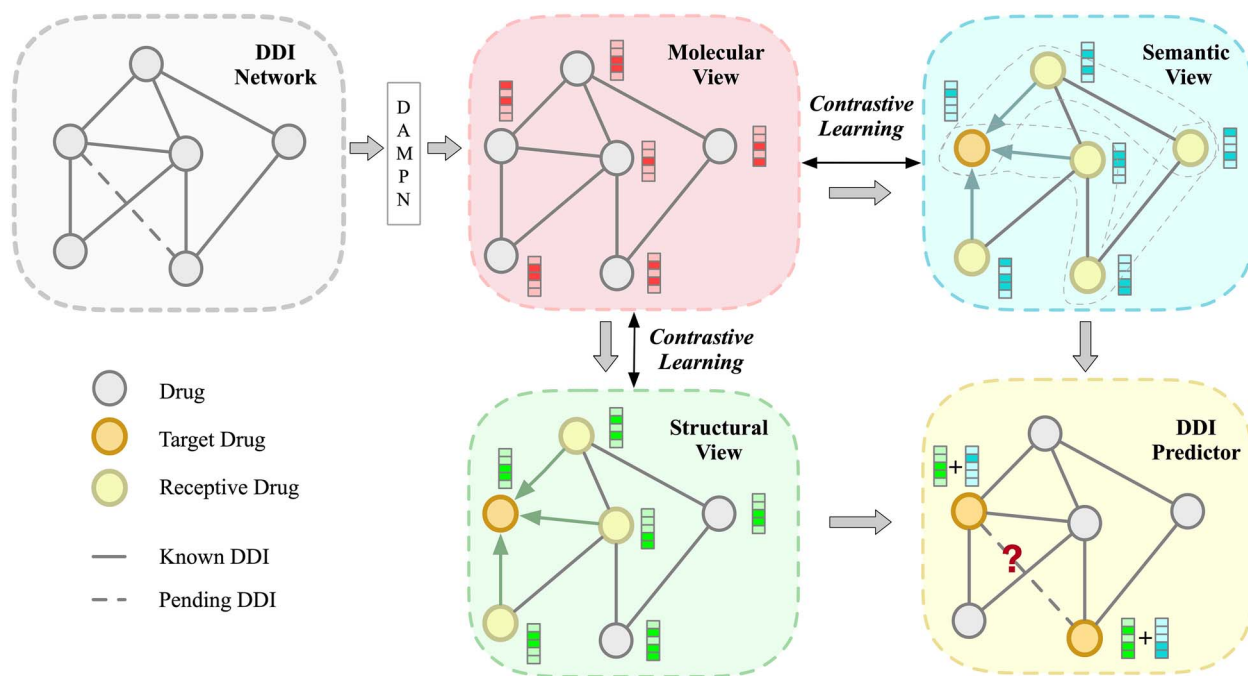
Figure 1. The overall architecture of HTCL-DDI. HTCL-DDI contains four hierarchical and interdependent modules: (A) In the molecular view, drug molecules are encoded into drug embeddings by a dual attention-aware message passing network; (B) In the structural view, the one-hop structural information is captured and integrated; (C) In the semantic view, the high-order semantic information and topological correlation are preserved by drug embedding propagation; (D) DDI predictor is designed to identify pending DDIs.

message is then propagated from each atom to its neighboring atoms, where the message is iteratively aggregated into the hidden representations of the neighbors. At each message passing layer, DAMPN concatenates the current atom representation with the hidden representation of the previous layer and then applies a linear transformation and a nonlinear function to learn a new hidden atom representation.

Specifically, we pass messages via chemical bonds and update the representations of atoms as follows:

$$\widetilde{a}_i^{(d)} = \sum_{t \in C_i} W_{b_{it}}^{(d)} \cdot a_t^{(d-1)} \qquad (1)$$

$$a_i^{(d)} = \sigma \left(W_a^{(d)} \cdot [a_i^{(d-1)}; \widetilde{a}_i^{(d)}] \right) \qquad (2)$$

where $\widetilde{a}_i^{(d)}$ and $a_i^{(d)}$ represent the candidate hidden embedding and the hidden embedding at depth d (D in total) of atom i, respectively, $W_{b_{it}}^{(d)}$ denotes the matrix of trainable parameters to reflect the message produced by bond b_{it} that connects atom i and atom t at depth d, C_i is the set of the neighboring atoms (including atom i) that connect atom i via the chemical bonds, $\sigma(\cdot)$ denotes the activation function and $[;]$ means the concatenation operation of vectors.

After D iterations of propagation, all of the atom embeddings within the given molecule are aggregated using an attention mechanism, generating a weighted combination representation denoted as h_{atom}. Formally speaking, the process can be defined as

$$e_i = \sigma \left(W_a \cdot [a_i^{(0)}; a_i^{(D)}] + b_a \right) \qquad (3)$$

$$h_{atom} = \sum_{i \in M_{atom}} e_i \cdot a_i^{(D)} \qquad (4)$$

where e_i denotes the trainable attention weight of atom i to indicate the importance of atom i, and M_{atom} denotes the set of atoms within the given molecule.

Chemical bonds prompt the exchange and transmission of information between atoms, and they are also closely related to the pharmacological properties of molecules. Similarly, we employ a linear layer to process chemical bond features and apply an attention mechanism to aggregate all chemical bonds within the given molecule, thereby obtaining an attention-aware chemical bond combination vector h_{bond}. The process can be formulated as follows:

$$f_j = \sigma \left(W_b \cdot b_j + b_b \right) \qquad (5)$$

$$h_{bond} = \sum_{j \in M_{bond}} f_j \cdot b_j \qquad (6)$$

where f_j denotes the trainable attention weight of bond j to reflect its importance, and M_{bond} denotes the set of bonds within the given molecule.

To learn a comprehensive representation h, we concatenate the representations of the atom combination vector h_{atom} and bond combination vector h_{bond} at the molecular level as follows:

$$h = [h_{atom}; h_{bond}] \qquad (7)$$

Structural view

After applying DAMPN introduced above, we move on to obtain the low-dimensional drug representations containing the intermolecular information. To model the structural information of the DDI network, we resort to a single-layer GAT [43] to aggregate

the one-hop neighbors, namely direct interactions in the DDI network.

The DDI network can be conceptualized as a graph with nodes representing drugs and edges representing DDIs. Some DDIs are clinically significant and pose serious risks to patients. Furthermore, certain DDIs reveal pivotal chemical reactions, which contribute valuable insights into DDI prediction with similar chemical substructures. Consequently, distinguishing the importance of each DDI is crucial for comprehending the overall impact of DDIs and predicting potential adverse effects.

The attention mechanism selectively emphasizes important drugs and suppresses irrelevant ones in the DDI network during the process of updating drug embeddings. Graph attention module assigns importance score α_{ij} to neighboring drug j of the given drug i as follows:

$$\alpha_{ij} = softmax\left(attention_{str}(h_i; h_j)\right) \tag{8}$$

where h_i and h_j represent the molecular representations of drug i and drug j in the molecular view, respectively.

To selectively concentrate on different parts of the DDI network, we extend GAT to multi-head GAT for stable training. The multi-head attention mechanism learns the multiple representations of drugs, each of which focuses on diverse aspects of the DDI network. By concatenating these representations, the structural view can be more sensitive to the complex drug associations in the DDI network. The drug embedding updated by the multi-head attention mechanism can be formally expressed as the following function:

$$d_{str(i)} = \overset{C}{\underset{c=1}{\Vert}} \sigma\left(\sum_{j \in N_i} [\alpha_{ij}]_c \cdot h_j\right) \tag{9}$$

where $\Vert$ is the concatenation operation, $\sigma(\cdot)$ denotes the activation function and $[\alpha_{ij}]_c$ indicates the normalized importance of neighboring drug j to given drug i at attention head c (C in total).

Semantic view

The structural view solely concentrates on one-hop neighbors directly connected to a given drug, which is far from enough. Furthermore, it is of vital importance to exploit the rich semantic information between drugs that exist beyond the immediate neighborhood for DDI prediction.

The single-layer GAT shows great strength in capturing node dependencies by aggregating one-hop neighbors, but the attention scores are mainly calculated based on node features, while the graph topology is only used to mask attention, resulting in the obtained embeddings lack of graph structure details. Previous studies have revealed that the performance of GAT decreases with increasing layers due to over-smoothing [44], reflecting the inherent limitation of GAT in preserving graph topological information. As depicted in Figure 2, we establish a three-layer GAT with topological information (TopoGAT) to discover high-order dependencies and effectively capture the latent semantic features in the DDI network. TopoGAT aggregates neighbors by both feature closeness and topological correlation, leading to a more comprehensive understanding of the DDI network.

The key aspect of TopoGAT lies in the computation of the attention score $\lambda_{ij}^{(l)}$. To compensate for the inherent limitations of GAT, we introduce a novel topological correlation term γ_{ij} into our model. Specifically, γ_{ij} is computed as the normalized count of all paths between drug i and drug j that can be traversed within
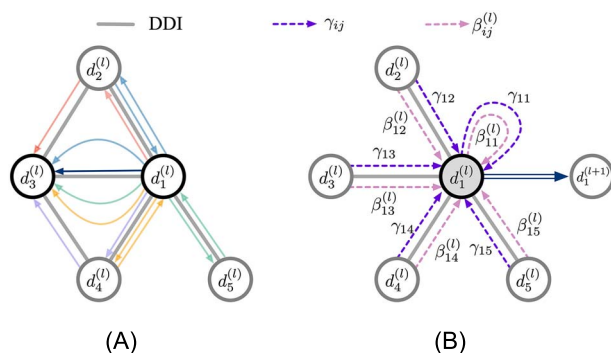
Figure 2. Explanation of TopoGAT. (A) A toy example of computing γ_{13}. Lines in different colors represent different pathways between drug 1 and drug 3 within a fixed number of hops; (B) The aggregation process of single-layer TopoGAT, including both feature closeness $\beta_{ij}^{(l)}$ and topological correlation γ_{ij}.

a fixed number of hops. γ_{ij} stores the topological correlation between drug i and drug j in the DDI network, with higher values tending toward more similar detailed graph structures.

$$\beta_{ij}^{(l)} = softmax\left(attention_{sem}(d_i^{(l-1)}; d_j^{(l-1)})\right) \tag{10}$$

$$\lambda_{ij}^{(l)} = softmax\left(w_\beta \cdot \beta_{ij}^{(l)} + w_\gamma \cdot \gamma_{ij}\right) \tag{11}$$

where $\beta_{ij}^{(l)}$ describes the normalized weight coefficient of GAT at layer l (L in total), w_β and w_γ are trainable parameters to balance feature information and topological structure.

To tackle the high variance of training, we apply the multi-head attention mechanism similar to the structural view. In the semantic view, we integrate high-order neighboring drugs to update the drug embedding $d_{sem(i)}$ as follows:

$$d_{sem(i)} = \overset{C}{\underset{c=1}{\Vert}} \sigma\left(\sum_{j \in N_i} [\lambda_{ij}^{(L)}]_c \cdot d_j^{(L-1)}\right) \tag{12}$$

DDI prediction
DDI predictor

We learn three drug representations (h, d_{str} and d_{sem}) of drug i and drug j through three hierarchical views. To investigate the potential interaction between drug pairs (i, j), we apply a fully connected neural network (FCN) to compress two drug embeddings into an interaction link vector P_{ij}, indicating the probability of interaction between drug i and drug j. According to the DDI predictor, three prediction scores are yielded to better optimize the model during training as follows:

$$P_{ij}^h = \sigma[FCN(h_i \odot h_j)] \tag{13}$$

$$P_{ij}^{d_{str}} = \sigma[FCN(d_{str(i)} \odot d_{str(j)})] \tag{14}$$

$$P_{ij}^{d_{sem}} = \sigma[FCN(d_{sem(i)} \odot d_{sem(j)})] \tag{15}$$

where $FCN(\cdot)$ indicates the FCN with two hidden layers, and $\odot$ denotes Hadamard product.

It is worth mentioning that we perform vector addition between the drug embeddings obtained from the structural and semantic views to generate the final drug representation for DDI prediction in the inference stage.

Loss function

The loss comprises three distinct components: the supervised loss, the inconsistency loss and the contrastive loss.

Supervised loss The supervised loss $\mathcal{L}_S$ describes the distance between the prediction results and the ground truth. In particular, we calculate and sum the distance between the predicted results obtained from three views and the true labels. The supervised loss $\mathcal{L}_S$ can be mathematically expressed as

$$\mathcal{L}_S = \sum_{(i,j)\in\Omega} BCE\left(P_{ij}^h \parallel L_{ij}\right) + BCE\left(P_{ij}^{d_{str}} \parallel L_{ij}\right) \tag{16}$$
$$+ BCE\left(P_{ij}^{d_{sem}} \parallel L_{ij}\right)$$

where (i,j) represents a drug pair in the training set, Ω denotes the set of observed drug pairs in the training set, L_{ij} denotes the true label of drug pair (i,j) and *BCE* means Binary Cross-Entropy.

Inconsistency loss The inconsistency loss $\mathcal{L}_I$ quantifies the dissimilarity among three prediction results of the unobserved samples. The purpose of minimizing the inconsistency loss is to prompt HTCL-DDI to improve the commonality and consistency between any two different views in three views.

$$\mathcal{L}_I = \sum_{(i',j')\in\Omega^-} KL\left(P_{i'j'}^{d_{str}} \parallel P_{i'j'}^h\right) + KL\left(P_{i'j'}^{d_{sem}} \parallel P_{i'j'}^h\right) \tag{17}$$
$$+ KL\left(P_{i'j'}^{d_{sem}} \parallel P_{i'j'}^{d_{str}}\right)$$

where (i',j') represents a drug pair not seen in the training set, Ω^- (the complement of Ω) denotes the set of unobserved drug pairs during training and *KL* means Kullback–Leibler divergence.

Contrastive loss To construct the contrastive loss, we develop a mutual information (MI) estimator on different view pairs to maximize the estimated MI over the given dataset. We select positive drug pairs and negative drug pairs and utilize a contrastive objective to ensure the representations of positive samples are consistent, while the representations of negative samples are distinguished. To address the challenge of estimating MI between high-dimensional representations, we introduce Jensen–Shannon divergence (JSD) [45], a measure of the difference between two probability distributions, which can be computed efficiently via neural networks. The MI between the intra-molecular level and the inter-molecular level is important. Therefore, we use $\mathcal{L}_{C_1}$ and $\mathcal{L}_{C_2}$ to measure MI between the structural view and the molecular view, and MI between the semantic view and the molecular view, respectively:

$$\mathcal{L}_{C_1} = \sum_{i\in\mathcal{N}} \frac{1}{|N_i|} \sum_{j\in N_i} \mathbb{E}_{\mathbb{D}\times\tilde{\mathbb{D}}}\left[sp\left(\tau_{\omega,\phi}\left(h_j^\omega(x'); \, d_{str(i)}^\phi(x)\right)\right)\right] \tag{18}$$
$$- \mathbb{E}_{\mathbb{D}}\left[- sp\left(-\tau_{\omega,\phi}\left(h_j^\omega(x); \, d_{str(i)}^\phi(x)\right)\right)\right]$$

$$\mathcal{L}_{C_2} = \sum_{i\in\mathcal{N}} \frac{1}{|N_i|} \sum_{j\in N_i} \mathbb{E}_{\mathbb{D}\times\tilde{\mathbb{D}}}\left[sp\left(\tau_{\omega,\psi}\left(h_j^\omega(x'); \, d_{sem(i)}^\psi(x)\right)\right)\right] \tag{19}$$
$$- \mathbb{E}_{\mathbb{D}}\left[- sp\left(-\tau_{\omega,\psi}\left(h_j^\omega(x); \, d_{sem(i)}^\psi(x)\right)\right)\right]$$

where N_i denotes the set of one-hop neighbors of drug i including itself, $\mathbb{D}$ means an empirical probability distribution, $sp(\cdot)$ is the softplus function, ω, ϕ and ψ denote the sets of parameters of DAMPN, GAT and TopoGAT, respectively.

The contrastive loss $\mathcal{L}_C$ can be defined as the sum of $\mathcal{L}_{C_1}$ and $\mathcal{L}_{C_2}$ as follows:

$$\mathcal{L}_C = \mathcal{L}_{C_1} + \mathcal{L}_{C_2} \tag{20}$$

Total loss Combined with the supervised loss $\mathcal{L}_S$, the inconsistency loss $\mathcal{L}_D$ and the contrastive loss $\mathcal{L}_C$, the total loss $\mathfrak{L}$ of HTCL-DDI can be formulated as follows:

$$\mathfrak{L} = \lambda_1\mathcal{L}_S + \lambda_2\mathcal{L}_I + \lambda_3\mathcal{L}_C \tag{21}$$

where λ_1, λ_2 and λ_3 are hyper-parameters to balance the three loss components. By minimizing $\mathfrak{L}$, HTCL-DDI is solved and optimized via the back-propagation algorithm.

RESULTS
Experiment preparation

To evaluate the scalability and robustness of HTCL-DDI, we test our model on three public datasets, which vary in scale and density. The scale of the dataset is determined by the number of drugs included. Due to some nonstandard SMILES strings, we remove the abnormal data, resulting in a certain reduction in the scale of datasets. Then we treat the observed DDIs as positive samples and randomly sample the non-existing DDIs to generate the negative samples. In addition, we divide the training set and testing set at a ratio of approximately 4:1, with 25% of the training set randomly selected to serve as the validation set. The statistics of the preprocessed datasets are listed as follows:

- ZhangDDI dataset[46] is of small-scale, consisting of 544 drugs and 45,720 pairwise DDIs.
- ChCh-Miner dataset[47] is of medium-scale, consisting of 997 drugs and 21,486 pairwise DDIs.
- DeepDDI dataset[20] is of large-scale, consisting of 1,704 drugs and 191,870 pairwise DDIs.

DDI prediction can be regarded as a binary classification task, where the goal is to predict whether a pair of drugs will interact or not. In our study, we utilize four widely used evaluation metrics to comprehensively assess the performance of DDI prediction methods. Specifically, we select the area under the receiver operating characteristic curve (AUROC), average precision (AP), F1-score (F1) and accuracy (ACC) as our evaluation metrics. To reduce the impact of random errors and draw reliable conclusions, we repeat all the experiments five times and report *mean ± standard deviation* of the four metrics.

In comparative experiments, we compare nine state-of-the-art methods: MR-GNN, GCN-BMP, EPGCN-DS, DeepDrug, MIRACLE, SSI-DDI, CSGNN, DeepDDS and DSN-DDI as follows:

- MR-GNN[48] adopts the multi-resolution architecture to capture local features of each graph and extract the interaction features between pairwise graphs.
- GCN-BMP[30] applies GNN in DDI prediction and utilizes an end-to-end graph representation learning framework for DDI prediction.
- EPGCN-DS[23] identifies DDIs from molecular structures and contains an encoder with expressive GCN layers and a decoder that outputs the DDI possibility.
- DeepDrug[22] uses residual graph convolutional networks (RGCNs) and convolutional networks (CNNs) to boost DDI prediction accuracy.

Table 1. Overall performance comparisons of different DDI prediction methods. The best results are highlighted in **bold**, and the runner-up results are highlighted in underline. (Higher values indicate better performance)

Dataset	Metric	MR-GNN	GCN-BMP	EPGCN-DS	DeepDrug	MIRACLE	SSI-DDI	CSGNN	DeepDDS	DSN-DDI	HTCL-DDI
ZhangDDI	AUROC	96.18±0.25	84.42±1.21	90.83±0.66	93.35±0.20	96.44±0.35	93.14±0.29	91.71±0.09	93.20±0.23	91.13±0.15	**98.58**±0.21
	AP	92.63±0.30	80.20±1.57	88.96±0.88	92.33±0.23	93.09±0.53	92.09±0.73	92.08±0.31	86.42±0.30	**97.06**±0.38
	F1	82.93±0.81	71.86±2.71	80.07±0.86	82.89±0.27	85.16±0.27	81.96±1.24	83.60±0.73	82.79±0.42	87.68±0.40	**92.19**±0.56
	ACC	91.90±0.50	75.78±1.07	82.40±1.04	85.67±0.33	93.16±0.16	85.35±0.50	84.14±0.45	85.63±0.28	86.65±0.46	**96.59**±0.24
ChCh-Miner	AUROC	93.11±0.36	78.65±0.56	94.23±0.71	98.38±0.10	96.20±0.79	98.09±0.14	97.68±0.10	97.10±0.18	96.69±0.20	**99.06**±0.15
	AP	95.95±0.19	86.31±0.54	96.80±0.40	99.16±0.05	99.50±0.11	98.97±0.06	97.56±0.16	98.51±0.08	96.34±0.27	**99.87**±0.02
	F1	88.13±0.72	80.87±0.92	89.41±0.66	94.67±0.26	94.55±0.66	93.98±0.34	92.47±0.22	92.21±0.63	88.12±0.64	**97.48**±0.19
	ACC	85.03±0.62	73.07±0.80	86.64±0.98	93.18±0.35	90.77±0.11	92.19±0.48	92.54±0.17	90.38±0.64	88.89±0.42	**95.61**±0.32
DeepDDI	AUROC	93.35±0.17	77.19±0.63	85.93±0.24	91.74±0.14	92.76±0.38	91.79±0.48	94.01±0.25	94.38±0.63	93.22±0.10	**94.49**±0.20
	AP	94.56±0.09	81.70±0.60	88.72±0.12	92.99±0.18	96.77±0.18	93.47±0.44	94.17±0.30	95.68±0.56	92.87±0.15	**97.41**±0.10
	F1	90.07±0.49	80.10±0.26	84.86±0.38	89.39±0.09	93.54±0.70	88.23±0.49	86.01±0.63	91.27±0.54	85.60±0.15	**94.78**±0.27
	ACC	87.54±0.43	72.94±0.49	80.22±0.39	86.28±0.12	90.33±0.98	85.38±0.59	86.33±0.36	88.87±0.68	85.41±0.09	**92.08**±0.37

- MIRACLE [33] provides a multi-view framework to capture intra-view molecular structure and inter-view DDIs between molecules simultaneously.
- SSI-DDI [31] breaks down the DDI prediction task between two drugs to identify pairwise interactions between their respective substructures.
- CSGNN [35] injects a mix-hop neighborhood aggregator into a GNN to capture high-order dependency in DDI networks and leverages a contrastive self-supervised learning task as a regularizer.
- DeepDDS [27] is a deep learning model based on GNNs and attention mechanism to identify the synergistic drug combinations.
- DSN-DDI [37] is a dual-view drug representation learning network to learn drug substructures from the single drug and the drug pair simultaneously.

We conduct all experiments on a CentOS release 7.9 system and employ the NVIDIA A100-PCIE GPU card with 40 GB of memory. To ensure a fair performance comparison, all the models are implemented in PyTorch. We set $\lambda_1 = 1.0$, $\lambda_2 = 0.8$, $\lambda_3 = 1.0$, and fix the number of epochs to 100. In addition, the depth of DAMPN and the number of TopoGAT layers are both set to 3, and the number of attention heads in GAT and TopoGAT are both set to 8. We initialize HTCL-DDI by Xavier [49] initialization and use Adam [50] optimizer to update the parameters of the entire model.

Overall performance

Table 1 lists the experimental results of the proposed HTCL-DDI and other baseline methods on three datasets. All the results manifest the superiority of HTCL-DDI, which achieves improved performance on multiple evaluation metrics across multiple datasets.

As a multi-view contrastive learning framework, HTCL-DDI outperforms all the single-view baseline methods, demonstrating the effectiveness of applying multi-view contrastive learning to DDI prediction. HTCL-DDI treats the same drug from diverse perspectives and leverages contrastive learning to enable the exchange and transmission of discriminative features across multiple views for more reliable and versatile drug representations. In addition, HTCL-DDI surpasses other multi-view contrastive learning frameworks on three datasets, demonstrating HTCL-DDI extends the potential of multi-view contrastive learning in DDI

prediction and facilitates to capture critical features relevant to DDIs.

Ablation study

The superior performance of HTCL-DDI largely benefits from the following four ingenious designs: the introduction of triple-view contrastive learning into DDI prediction, the combination of multiple loss components, the incorporation of bond information in the molecular view and the utilization of TopoGAT in the semantic view. To further investigate their contributions to experimental performance, we conduct ablation studies on our hierarchical triple-view contrastive learning framework and specially tailored DAMPN and TopoGAT.

Ablation study on triple views

The multi-view contrastive learning architecture serves as the fundamental basis for HTCL-DDI. To explore the effects of integrating multiple views on DDI prediction performance, we conduct experiments utilizing the single view and the dual views on ZhangDDI dataset.

As illustrated in Table 2, the introduction of the structural view significantly enhances the DDI prediction performance, and the semantic view also provides useful information for DDI prediction. These experimental findings effectively illuminate that the incorporation of multiple views enables HTCL-DDI to perceive multi-level information, which sheds light on a promising direction for DDI prediction. The multi-view framework offers great potential to tackle the issues of the single-view paradigm in dealing with complex data structures. In particular, single-view methods that solely rely on molecular structures may disregard the contextual information in DDI networks. Furthermore, multi-view methods possess the ability to capture both the fine-grained details and the overall patterns of DDIs and also facilitate mutual enhancement across multiple views via contrastive learning, thereby providing comprehensive and informative drug representations.

Ablation study on multiple loss components

In addition to the supervised loss $\mathcal{L}_S$ based on the true labels, we incorporate two unsupervised loss components: the inconsistency loss $\mathcal{L}_I$ and the contrastive loss $\mathcal{L}_C$. We conduct ablation experiments to individually remove these two loss components to assess the impacts of these two unsupervised losses on model performance.

Table 2. The experimental results for ablation study of triple views on ZhangDDI dataset. (The best results are highlighted in **bold**.)

Metric	single view[1]	dual views[2]	triple views
AUROC	92.01±1.02	96.18±0.25	**98.58**±0.21
AP	85.34±1.24	92.63±0.30	**97.06**±0.38
F1	76.58±0.59	82.94±0.81	**92.19**±0.56
ACC	88.57±0.24	91.90±0.50	**96.59**±0.24

[1]: the molecular view. [2]: the molecular view and the structural view.

Table 3. The experimental results for ablation study of multiple loss components on ZhangDDI dataset. (The best results are highlighted in **bold**.)

Metric	w/o $\mathcal{L}_I$	w/o $\mathcal{L}_C$	w/ $\mathcal{L}_I$&$\mathcal{L}_C$
AUROC	97.84±0.24	98.13±0.15	**98.58**±0.21
AP	95.59±0.34	96.14±0.23	**97.06**±0.38
F1	88.74±0.28	88.43±0.35	**92.19**±0.56
ACC	94.62±0.31	94.66±0.21	**96.59**±0.24

Table 4. The experimental results for ablation study of DAMPN on ZhangDDI dataset. (The best results are highlighted in **bold**.)

Metric	SAMPN	DAMPN
AUROC	98.44±0.15	**98.58**±0.21
AP	96.71±0.25	**97.06**±0.38
F1	89.98±0.38	**92.19**±0.56
ACC	95.39±0.23	**96.59**±0.24

According to Table 3, the introduction of unsupervised loss contributes to DDI prediction performance. Specifically, the inconsistency loss encourages the multiple views to converge toward the same DDI prediction, thereby enhancing the robustness of HTCL-DDI. The contrastive loss encourages HTCL-DDI to focus on relevant features and discard irrelevant ones, leading to discriminative and informative representations of drugs.

Ablation study on DAMPN

DAMPN provides an effective tool for acquiring compositional and structural information of drug molecules, which simultaneously aggregates atoms and bonds to fully exploit intra-molecular information. To evaluate the impact of combining chemical bond information, we perform ablation experiments by attentively aggregating the atom embeddings and ignoring the bond embeddings to generate the drug representations at the intra-molecular level.

As listed in Table 4, we substitute DAMPN with single attention-aware message passing network (SAMPN) in the molecular view. Experimental results manifest that combining the bond information can enhance DDI prediction performance, as chemical bonds not only facilitate message transmission between the connected atoms but also contain useful information for DDI prediction.

Ablation study on TopoGAT

TopoGAT is an essential element of capturing rich semantic information in the semantic view. We improve upon the classic GAT architecture to design TopoGAT. Specifically, TopoGAT incorporates the topological details that the number of pathways reachable within k hops between any drug pairs into GAT architecture. As reported in Table 5, additional experiments are performed by substituting TopoGAT with GAT and GCN in the semantic view.

The experimental results support that TopoGAT possesses a powerful capacity for handling semantic information in the

Table 5. The experimental results for ablation study of TopoGAT on ZhangDDI dataset. (The best results are highlighted in **bold**.)

Metric	GCN	GAT	TopoGAT
AUROC	97.16±0.47	98.16±0.12	**98.58**±0.21
AP	94.00±0.72	96.46±0.15	**97.06**±0.38
F1	84.57±0.38	88.41±0.48	**92.19**±0.56
ACC	92.74±0.68	94.52±0.46	**96.59**±0.24

complicated DDI network compared with GAT and GCN. The introduction of both feature information and topological dependencies when updating attention scores is implied to contribute to informative drug representations. TopoGAT facilitates the interpretation of the learned representations by unearthing the underlying topological properties of DDI networks. The additional information enhances the ability to distinguish between drugs with similar molecular structures but different network topologies, leading to the improved DDI prediction performance.
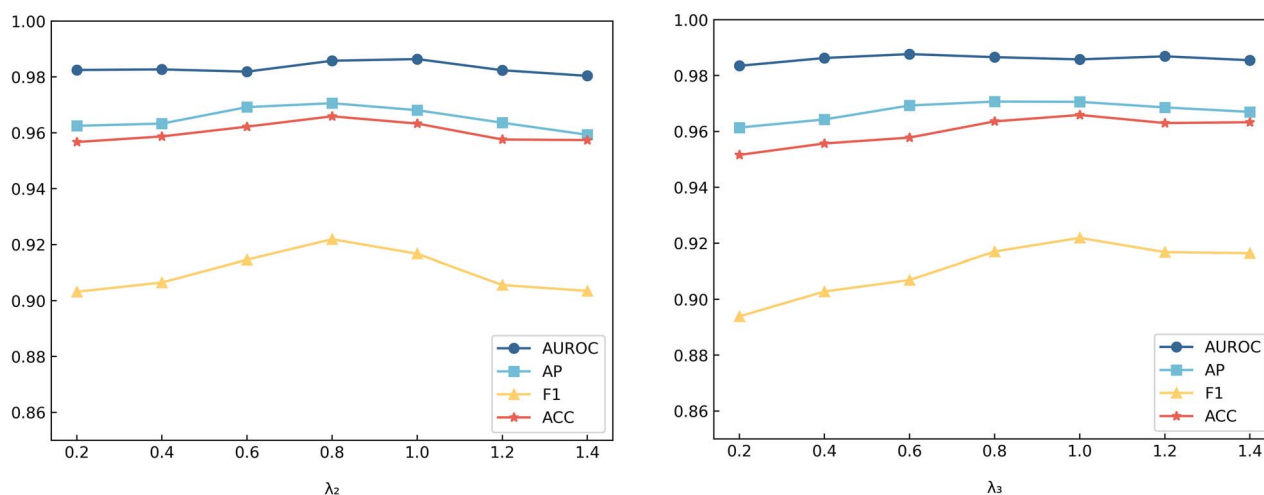
Parameter sensitivity

Sensitivity analysis is crucial for understanding how changes in hyperparameters affect HTCl-DDI's performance. In particular, λ_1, λ_2 and λ_3 balance the supervised loss $\mathcal{L}_S$, the inconsistency loss $\mathcal{L}_I$ and the contrastive loss $\mathcal{L}_C$, respectively. Therefore, we fix $\lambda_1 = 1$, and vary λ_2 and λ_3 by {0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4} to find their optimal values. According to Figure 3, the near optimal comprehensive performance at $\lambda_2 = 0.8$ and $\lambda_3 = 1.0$ justifies our parameter settings.
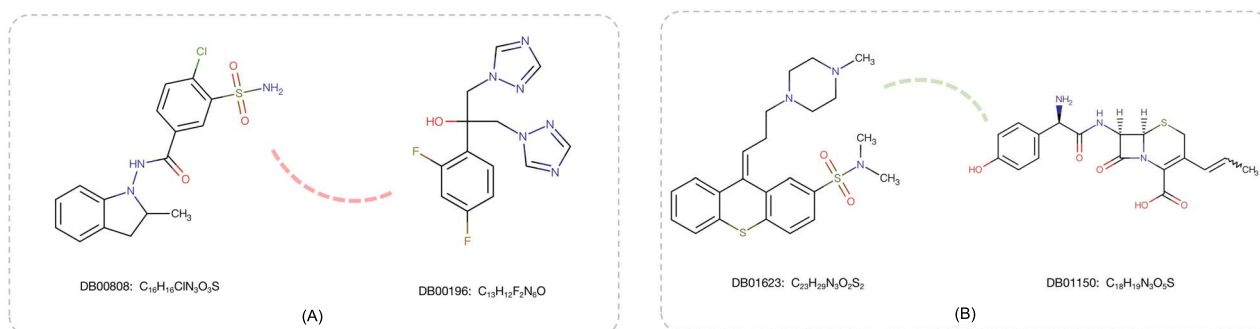
Case study

To verify the application potential of HTCL-DDI in real-world scenarios, we analyze clinical studies on HTCL-DDI's prediction results of two drug pairs in ZhangDDI testing set in Figure 4. Specifically, (1) Indapamide and Fluconazole can potentially interact with each other, which is consistent with the prediction of HTCL-DDI. Fluconazole is an antifungal medication that inhibits the activity of certain enzymes in the liver, which play a role in the metabolism of Indapamide. Therefore, taking Fluconazole with Indapamide is prone to an increase in the blood levels of Indapamide, potentially causing an increased risk of side effects. (2) According to the inference of HTCL-DDI, no interaction between Thiothixene and Cefprozil will occur. In fact, Thiothixene and Cefprozil have different absorption, distribution, metabolism and excretion profiles. Existing literature and clinical studies have not reported any significant adverse effects or pharmacological interactions. The two cases reflect that HTCL-DDI can provide an effective tool for ensuring drug safety and identifying synergistic drug combinations.

CONCLUSION

In this paper, we present a hierarchical perspective to reformulate the DDI prediction into an end-to-end computational link prediction task and then construct a two-tier framework that consists of the molecular view, structural view and semantic view. The goal of the molecular view is to exploit the intra-molecular information, including the internal composition and molecular structure. As for the interactions between molecules, the structural view aims to preserve the one-hop structural information, and the semantic view intends to utilize the high-order semantic information. Specifically, we adopt DAMPN to aggregate atoms and bonds via

Figure 3. Parameter sensitivity study of λ_2 and λ_3 on ZhangDDI dataset.

Figure 4. Case study of DDI prediction results: (A) Indapamide and Fluconazole tend to interact with each other; (B) Thiothixene and Cefprozil are less likely to interact with each other.

a dual-attention mechanism at the intra-molecular level and establish a single-layer multi-head GAT and a three-layer multi-head TopoGAT to capture structural and semantic information at the inter-molecular level. We also introduce contrastive learning across triple views to learn informative drug representations and further design a DDI predictor to distinguish unknown DDIs. Extensive experiments on three public datasets evaluate the performance and examine the contributions of different modules in HTCL-DDI. All the empirical results demonstrate that HTCL-DDI can provide an effective and promising tool for DDI prediction, which can support medication safety and drug side effect research. Our work can be further improved from the following three aspects: (1) The heterogeneous biomedical information can be incorporated to enhance representation learning; (2) HTCL-DDI is expected to be extended to more complex and practical application scenarios; (3) The wet-lab experiments can be supplemented to further verify part of the DDI prediction results.

Key Points

- We construct a two-tier framework with triple views to describe complicated DDI patterns and capture hierarchical information from different perspectives including the molecular, structural and semantic views. HTCL-DDI uses contrastive learning to facilitate mutual enhancement across three views and improve the capability to handle both local and global information.

- We design a novel dual-attention mechanism in the molecular view to generate informative molecular representations and introduce topological information into GAT in the semantic view to model high-order relations, fully exploiting topological correlation to mitigate the intrinsic over-smoothing problem in GAT.

- Extensive experiments conducted on three real-world datasets demonstrate that HTCL-DDI achieves significant performance improvements over the state-of-the-art DDI prediction methods.

DATA AVAILABILITY STATEMENT

The data and source code used in HTCL-DDI are available at: https://github.com/ranzhran/HTCL-DDI.

REFERENCES

1. Kaufman DW, Kelly JP, Rosenberg L, *et al.* Recent patterns of medication use in the ambulatory adult population of the United States: the slone survey. *JAMA* 2002;**287**(3):337–44.

2. Fialová D, Topinková E, Gambassi G, *et al*. Potentially inappropriate medication use among elderly home care patients in europe. *JAMA* 2005;**293**(11):1348–58.

3. Lai X, Zhu H, Huo X, Li Z. Polypharmacy in the oldest old (≥ 80 years of age) patients in China: a cross-sectional study. *BMC Geriatr* 2018;**18**(1):1–8.

4. Safdari R, Ferdousi R, Aziziheris K, *et al*. Computerized techniques pave the way for drug-drug interaction prediction and interpretation. *Bioimpacts* 2016;**6**:71–8.

5. Han K, Peigang Cao Y, Wang FX, *et al*. A review of approaches for predicting drug–drug interactions based on machine learning. *Front Pharmacol* 2022;**12**:3966.

6. Segura-Bedmar I, Fernández PM, Zazo MH. *Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013)*. Atlanta, Georgia, USA: Association for Computational Linguistics, 2013;**2**:341–350.

7. Pathak J, Kiefer RC, Chute CG. Using linked data for mining drug-drug interactions in electronic health records. *Stud Health Technol Inform* 2013;**192**:682–6.

8. Zhang P, Wang F, Jianying H, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**(1): 12339.

9. Gottlieb A, Stein GY, Oron Y, *et al*. Indi: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 2012;**8**(1): 592.

10. 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.

11. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform* 2017;**70**:54–64.

12. Vilar S, Uriarte E, Santana L, *et al*. Detection of drug-drug interactions by modeling interaction profile fingerprints. *PloS One* 2013;**8**(3): e58321.

13. Shi J-Y, Huang H, Li J-X, *et al*. Tmfuf: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs. *BMC Bioinformatics* 2018;**19**(14): 27–37.

14. Zhang W, Chen Y, Li D, Yue X. Manifold regularized matrix factorization for drug-drug interaction prediction. *J Biomed Inform* 2018;**88**:90–7.

15. Rohani N, Eslahchi C, Katanforoush A. Iscmf: integrated similarity-constrained matrix factorization for drug–drug interaction prediction. *Netw Model Anal Health Inform Bioinform* 2020;**9**: 1–8.

16. Hui Y, Mao K-T, Shi J-Y, *et al*. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol* 2018;**12**(1):101–10.

17. Cami A, Manzi S, Arnold A, Reis BY. Pharmacointeraction network models predict unknown drug-drug interactions. *PloS One* 2013;**8**(4): e61468.

18. Shahreza ML, Ghadiri N, Mousavi SR, *et al*. A review of network-based approaches to drug repositioning. *Brief Bioinform* 2018;**19**(5):878–92.

19. Cheng Yan, Duan Guihua, Zhang Yayan, Wu Fang-Xiang, Pan Yi, and Wang Jianxin. Idnddi: An integrated drug similarity network method for predicting drug-drug interactions. In Cai Z, Skums P, Li M. (eds) *Bioinformatics Research and Applications: 15th International Symposium, ISBRA 2019, Barcelona, Spain, June 3–6, 2019, Proceedings 15*, **11490**: pp. 89–99. Cham: Springer, 2019. https://doi.org/10.1007/978-3-030-20242-2_8.

20. 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.

21. Liu S, Huang Z, Qiu Y, *et al*. Structural network embedding using multi-modal deep auto-encoders for predicting drug-drug interactions. In: *2019 IEEE International conference on bioinformatics and biomedicine (BIBM)*. San Diego, CA, USA: IEEE, 2019, 445–50. https://doi.org/10.1109/BIBM47256.2019.8983337.

22. Cao X, Fan R, Zeng W. Deepdrug: a general graph-based deep learning framework for drug relation prediction. *biorxiv* 2020. https://doi.org/10.1101/2020.11.09.375626.

23. Sun M, Wang F, Elemento O, Zhou J. Structure-based drug-drug interaction detection via expressive graph convolutional networks and deep sets (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence* 2020;**34**:13927–8.

24. Feng Y-H, Zhang S-W, Shi J-Y. Dpddi: a deep predictor for drug-drug interactions. *BMC Bioinformatics* 2020;**21**(1):1–15.

25. Yi H-C, You Z-H, Huang D-S, Kwoh CK. Graph representation learning in bioinformatics: trends, methods and applications. *Brief Bioinform* 2022;**23**(1): bbab340.

26. Guo L, Lei X, Chen M, Pan Y. Msresg: using gae and residual gcn to predict drug–drug interactions based on multi-source drug features. *Interdisc Sci* 2023;**15**:171–188. https://doi.org/10.1007/s12539-023-00550-6.

27. Wang J, Liu X, Shen S, *et al*. Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**(1): bbab390.

28. Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: a survey. *Knowledge-Based Systems* 2018;**151**: 78–94.

29. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**(13): i457–66.

30. Chen X, Liu X, Ji W. Gcn-bmp: investigating graph representation learning for ddi prediction task. *Methods* 2020;**179**:47–54.

31. Nyamabo AK, Hui Y, Shi J-Y. Ssi–ddi: substructure–substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;**22**(6): bbab133.

32. He H, Chen G, Chen CY-C. 3dgt-ddi: 3d graph and text based neural network for drug–drug interaction prediction. *Brief Bioinform* 2022;**23**(3): bbac134.

33. 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.

34. Ma T, Xiao C, Zhou J, Wang F. Drug similarity integration through attentive multi-view graph auto-encoders arXiv preprint arXiv:1804.10850. 2018. https://doi.org/10.48550/arXiv.1804.10850.

35. Zhao C, Liu S, Huang F, *et al*. Csgnn: contrastive self-supervised graph neural network for molecular interaction prediction. *In IJCAI* 2021;3756–63. https://doi.org/10.24963/ijcai.2021/517.

36. Li Z, Zhu S, Shao B, *et al*. Multi-view substructure learning for drug-drug interaction prediction arXiv preprint arXiv:2203.14513. 2022. https://doi.org/10.48550/arXiv.2203.14513.

37. Li Z, Zhu S, Shao B, *et al*. Dsn-ddi: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Brief Bioinform* 2023;**24**(1):bbac597. https://doi.org/10.1093/bib/bbac597.

38. 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.

39. Wang J, Zhang S, Li R, *et al.* Multi-view feature representation and fusion for drug-drug interactions prediction. *BMC Bioinformatics* 2023;**24**(1):1–23.

40. 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.

41. Gilmer J, Schoenholz SS, Riley PF, *et al.* Neural message passing for quantum chemistry. In: *International conference on machine learning*. Sydney, Australia: PMLR, 2017;**70**: 1263–72.

42. Landrum G, *et al.* Rdkit: a software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* 2013;**8**.

43. Velickovic P, Cucurull G, Casanova A, *et al.* Graph attention networks. *Stat* 2017;**1050**(20):10–48550.

44. Zhang K,, Zhu Y, Wang J, Zhang J. Adaptive structural fingerprints for graph attention networks. In *International Conference on Learning Representations*. Addis Ababa, Ethiopia: ICLR 2020.

45. Fuglede B, Topsoe F. Jensen-shannon divergence and hilbert space embedding. In *International symposium on Information theory,* 2004. ISIT 2004. *Proceedings*., pp. 31. Chicago, IL, USA: IEEE, 2004. .https://doi.org/10.1109/ISIT.2004.1365067.

46. Zhang W, Chen Y, Liu F, *et al.* Biosnap datasets: Stanford biomedical network dataset collection. *BMC Bioinformatics* 2017;**18**:1–12.

47. Maheshwari SZ, Zitnik M, Sosič R, Leskovec J. Predicting potential drug-drug interactions by integrating chemical, *biological, phenotypic and network data. [n. d.]*, http://snap.stanford.edu/biodata, 2018.

48. Xu N, Wang P, Chen L, *et al.* Mr-gnn: multi-resolution and dual graph neural network for predicting structured entity interactions arXiv preprint arXiv:1905.09558. 2019. https://doi.org/10.48550/arXiv.1905.09558.

49. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, **9**; pp. 249–56. Chia La-guna Resort, Sardinia, Italy: JMLR Workshop and Conference Proceedings, 2010.

50. Kingma DP, Ba J. Adam: a method for stochastic optimization arXiv preprint arXiv:1412.6980. 2014. https://doi.org/10.48550/arXiv.1412.6980.