BMC Bioinformatics

RESEARCH Open Access

# HLN-DDI: hierarchical molecular representation learning with co-attention mechanism for drug-drug interaction prediction

Yue Luo[1], Lei Deng[1*] and Zhijian Huang[1]

*Correspondence:
leideng@csu.edu.cn

[1] School of Computer Science and Engineering, Central South University, Changsha, China

## Abstract

**Background:** Accurate identification of drug-drug interactions (DDIs) is critical in pharmacology, as DDIs can either enhance therapeutic efficacy or trigger adverse reactions when multiple medications are administered concurrently. Traditional methods for identifying DDIs are labor-intensive and time-consuming, prompting the development of computational alternatives. However, existing computational approaches frequently encounter challenges related to interpretability and struggle to effectively capture the complex, multi-level structures inherent in drug molecules. Specifically, they often fail to adequately analyze substructural components and neglect interactions across hierarchical structural levels, resulting in incomplete molecular representations.

**Results:** In this study, we propose a Hierarchical Learning Network with a co-attention mechanism tailored to molecular structure representation for predicting DDIs, named HLN-DDI. The proposed method advances existing approaches by explicitly encoding motif-level structures and capturing hierarchical molecular representations at atom-level, motif-level, and whole-molecule scales. These hierarchical representations are integrated using a co-attention mechanism and combined with interaction-type information to enhance predictive performance. Comprehensive evaluations demonstrate that HLN-DDI significantly outperforms state-of-the-art methods across multiple benchmark datasets, achieving over 98% accuracy under transductive scenarios and surpassing 99% on various evaluation metrics. Moreover, HLN-DDI achieves a notable accuracy improvement of 2.75% in predicting DDIs involving unseen drugs. Practical assessments with real-world DDI scenarios further validate the efficacy and utility of our proposed model.

**Conclusion:** By leveraging hierarchical molecular structures and employing a co-attention mechanism to effectively integrate multi-level representations, HLN-DDI generates comprehensive and precise drug representations, leading to substantially improved predictions of potential drug-drug interactions.

**Keywords:** Drug-drug interaction, Molecular graph learning, Co-attention mechanism

## Introduction

Drugs, defined as chemical substances with specific structural characteristics, exert biological effects upon administration to living organisms [1]. Pharmaceutical drugs, also known as medications or medicines, are chemical compounds employed to prevent or treat diseases [1]. Drug-drug interactions (DDIs) significantly contribute to adverse drug reactions (ADRs). While some DDIs can yield beneficial synergistic effects, they more commonly compromise drug efficacy, safety, or resistance profiles, particularly when multiple medications are used concurrently [2–4]. The combination of multiple medications often becomes necessary due to the complexity of certain diseases or the simultaneous management of multiple diseases [5, 6]. For instance, during the COVID-19 pandemic, patients with pre-existing conditions such as coronary heart disease and hypertension have required concurrent administration of their regular medications alongside COVID-19 treatments, increasing the risk of adverse clinical effects [7]. Therefore, a thorough evaluation of pharmacological effects prior to combined medication use is essential for ensuring efficacy and reducing adverse drug reactions [8]. Although clinical trials and in vitro experiments are conventional methods for identifying DDIs [5, 9], many potential DDIs remain undetected due to the high costs, extensive sample preparation, sophisticated equipment requirements, and lengthy timeframes associated with traditional experimental approaches [10, 11]. Consequently, the development of computational techniques for large-scale DDI prediction has become a crucial research objective.

Recent advancements in computational hardware and rapid expansion of drug-related data have promoted the extensive application of bioinformatics in DDI prediction [12]. Unlike conventional experimental methods, computational modeling allows the identification of potential interactions based on insights derived from existing validated datasets. Graph Neural Networks (GNNs) have shown notable efficacy in capturing complex molecular structures [13]. For instance, MR-GNN [14] employs GNNs to generate node-level embeddings and captures distinct drug substructural representations, subsequently processed by a recurrent neural network (RNN) for DDI prediction. Molormer [15] integrates both 2D and 3D structural data for accurate DDI predictions. Additionally, MHCADDI [16] utilizes a co-attention mechanism to fuse internal and external drug features for improved DDI predictions.

Various research efforts have employed diverse GNN structures to encode molecular substructures into informative embeddings, thereby facilitating accurate DDI predictions [17–21]. These methods primarily model drug molecules as interconnected networks of functional groups and chemical substructures, which collectively influence drug pharmacological properties [22]. However, traditional GNN-based methods often aggregate information with limited chemical relevance, resulting in chemically insignificant substructures. Moreover, the path length for interatomic information transfer depends on neighborhood radii, often empirically determined through experimentation, significantly impacting model performance, particularly for DDIs involving unknown drugs. Zhu et al. [23] introduced the Molecular Substructure Aware Network (MSAN), employing a transformer-based framework to assign atoms to predefined substructure queries. Nevertheless, the arbitrary assignment and fixed number of substructures compromise chemical meaningfulness and interpretability. Additionally, existing models

typically aggregate node representations at a single structural level (atom-level or substructure-level), limiting their capacity to capture comprehensive molecular information and interactions across hierarchical structural layers.

To address these challenges, we propose a novel hierarchical learning network employing molecular structures and a co-attention mechanism for DDI prediction, designated HLN-DDI. Molecular motifs-small, conserved substructures such as triangles and cliques-are essential components in modeling biochemical and social networks [24, 25]. Our approach leverages these motifs, identified using an enhanced version of the BRISC algorithm [26], to enrich molecular graphs, thereby facilitating representation extraction at atom-level, motif-level, and molecule-level. A co-attention mechanism [27] integrates these multi-level representations by evaluating interaction scores across different structural hierarchies, incorporating relationship types to derive potential DDI probabilities.

Experimental results on two benchmark datasets demonstrate that HLN-DDI consistently surpasses state-of-the-art methods in both transductive and inductive scenarios. The hierarchical structure combined with the co-attention mechanism significantly enhances prediction accuracy. Ablation studies and visualization analyses further underscore the importance and effectiveness of hierarchical representation learning. Ultimately, real-world application assessments confirm the practical value and reliability of the proposed HLN-DDI model for predicting DDIs.

## Methods

### Problem formulation

The DDI prediction task can be conceptualized as a binary classification task. Specifically, this problem is framed in terms of tuples $(d_i, r, d_j)$, where the goal is to ascertain the existence of an interaction between drugs $d_i$ and $d_j$ for a given DDI type $r$. The input dataset comprises SMILES [28] representing the two drugs and the ordinal number corresponding to the relationship type to be predicted. To process this data, we utilize the RDKIT [29] to convert the SMILES representations, which are initially in a 2D sequence format, into 3D molecular graphs $G = (V, E)$. Here, $V = \{v_1, v_2, \ldots, v_n\}$ indicates the set of atoms, and $E = \{e_1, e_2, \ldots, e_m\}$ indicates the set of bonds within the molecule.

### Model architecture

In this research, we present HLN-DDI, an innovative approach for estimating the likelihood of interaction in drug pairs $(d_i, r, d_j)$. The architecture of HLN-DDI is depicted in Fig. 1. The model operates through three primary stages:

#### *Motif decomposition and augmented molecular graph construction*

The model receives as input the molecular graphs $G_i, G_j$ corresponding to drugs $d_i, d_j$, respectively. Using drug $d_i$ as an illustrative example, we dissect the molecular graph $G_i = (V_i, E_i)$ into a sequence of motifs denoted as $V_i^m = \{V_{i,1}^m, V_{i,2}^m, ..., V_{i,k}^m\}$, where $k$ signifies the total motifs into which $d_i$ can be divided. The Breakdown of Recurrent Chemical Substructures (BRICS) method [26] is a prevalent technique for decomposing and recombining molecules. Its objective is to identify and leverage recurring chemical motifs within molecules to create novel molecular structures, thereby broadening the explored chemical landscape. However, BRICS operates based on a finite set of chemical

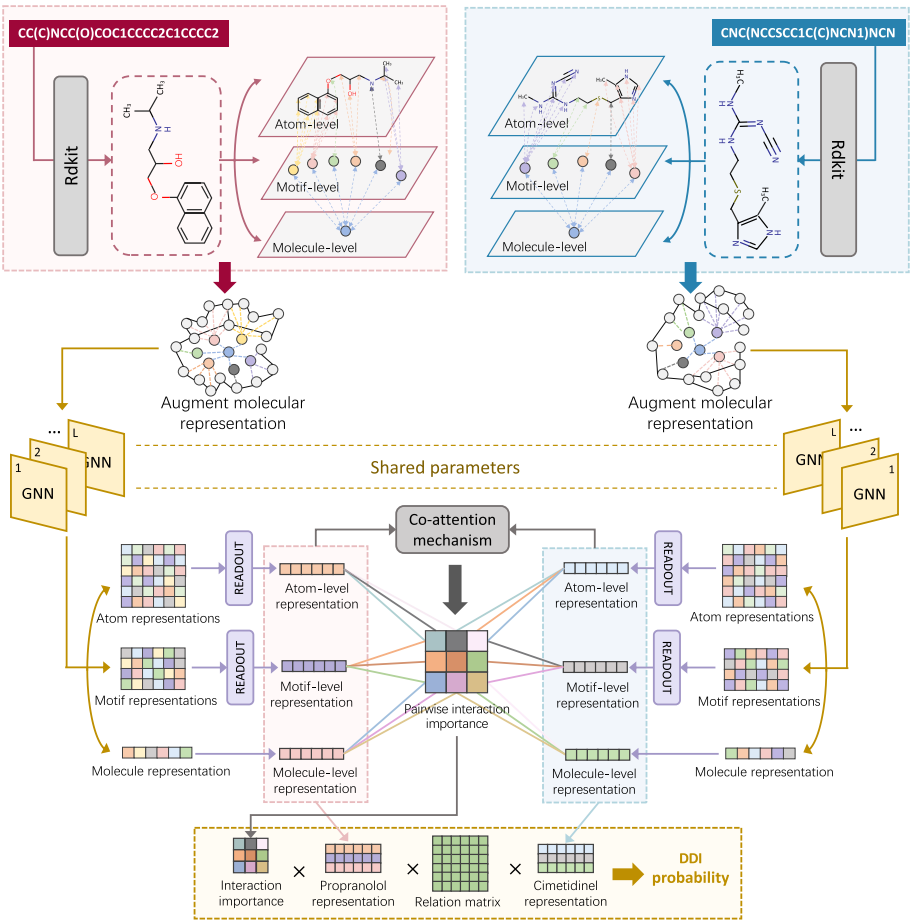


**Fig. 1** Framework of HLN-DDI. HLN-DDI first decomposes the input molecular graphs of two drugs into motifs, subsequently augmenting these graphs by incorporating motif-level and molecular-level nodes and establishing corresponding edges. Graph Neural Networks (GNNs) with shared weights are then employed to hierarchically encode the augmented graphs, generating atom-level, motif-level, and molecular-level representations. These hierarchical representations are integrated via a co-attention mechanism to compute a pairwise interaction importance matrix. Finally, the likelihood of drug-drug interactions is determined by weighting and aggregating the drug representations and relationship information based on the interaction importance matrix

rules, resulting in a relatively coarse decomposition granularity. Following the BRICS decomposition, some larger motifs may still contain more generic and functional substructures that require further segmentation. Inspired by the work of Zang et al. [30], we augment the foundational BRICS method with an additional decomposition principle: disintegrating larger ring fragments and selecting their smallest constituent ring as the resultant motif. The meticulous process of motif decomposition utilized in this study is illustrated in Fig. 2.

Utilizing the decomposition outcomes, we incorporated all derived motifs as nodes $V_i^m$ within $G_i$. Additionally, a molecule-level node $V_i^g$ was crafted for augmenting $G_i$. Subsequently, we established atom-motif edges $E_i^m$, which are obtained by linking each motif-level node to all the atom-level nodes that comprise it. Furthermore, we connected the molecule-level nodes to all the motif-level nodes in $G_i$ forming motif-molecular bonds

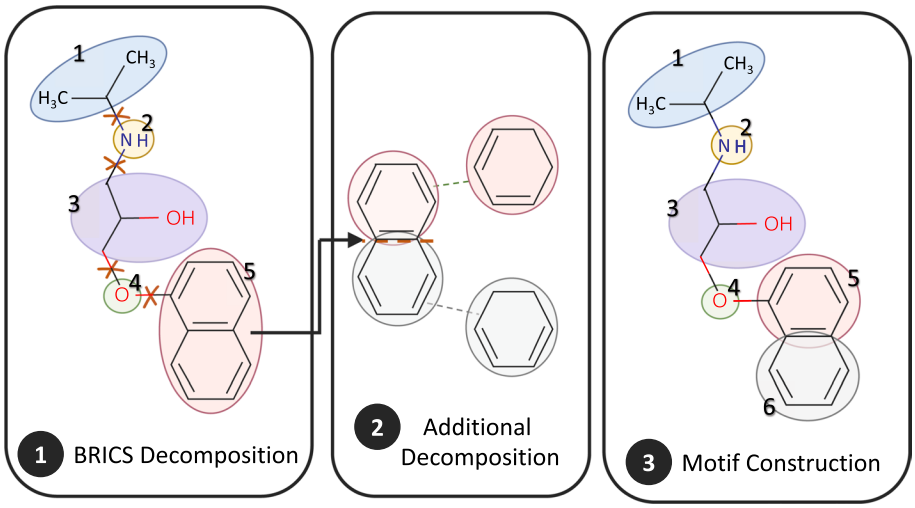


**Fig. 2** Summary of Motif Creation Process. The entire construction encompasses three sequential stages: **1** Initially, the provided molecular graph is broken down utilizing the BRICS framework, **2** Subsequently, the molecule undergoes additional decomposition guided by our supplementary rules, and **3** Finally, the resultant motifs are assembled into motif-level nodes

$E_i^g$. The detailed compositional components included in the final augmented molecular graph are given by Eq. 1.

$$\tilde{G}_i = (\tilde{V}_i, \tilde{E}_i), \tilde{V}_i = [V_i, V_i^m, V_i^g], \tilde{E}_i = [E_i, E_i^m, E_i^g] \tag{1}$$

Analogously, the augmented molecular graph $\tilde{G}_j$ corresponding to $d_j$ can be derived by following the identical procedures outlined previously.

#### *Hierarchical representation encoding and feature aggregation*

After constructing two augmented representations of drug molecular graphs, we utilize the GNN to encode the node representations at each hierarchical level. The GNN framework primarily relies on the Message Passing mechanism [31], which involves aggregating knowledge from the neighboring nodes and subsequently updating the feature of the central node. Various notable GNN architectures include the Graph Convolutional Network (GCN) [32], Graph Attention Network (GAT) [33], and Graph Isomorphism Network (GIN) [34]. In our framework, we adopt the GIN as the backbone GNN architecture for encoding molecular graphs. Compared to other commonly used architectures such as GCN and GAT, GIN offers stronger discriminative power in distinguishing different graph structures. This is supported by theoretical analysis from Xu et al. [34], which demonstrates that GIN is as powerful as the Weisfeiler-Lehman (WL) graph isomorphism test in distinguishing non-isomorphic graphs. This property is particularly important in molecular representation learning, where small changes in graph structure (e.g., local substructures or atom connectivity) can lead to significantly different chemical properties. Moreover, GIN has been successfully applied in various molecular machine learning tasks, such as property prediction (e.g., QM9 and ZINC) [34], drug-target interaction prediction [35], and toxicity classification [36]. Therefore, using GIN enhances our model's ability to capture subtle yet crucial structural patterns within

molecular graphs. Given these considerations, we opted for GIN with shared parameters to encode the two augmented molecular graphs of drugs, and the specific aggregation pattern for layer $l$ is detailed below.

$$h_v^l = MLP_a^l(h_v^{l-1} + \sum_{u \epsilon N(v)} (h_v^{l-1} + MLP_b^l(X_{uv}^b)) \tag{2}$$

In this context, $MLP_a^l$ and $MLP_b^l$ denote the l-th layer of MLPs responsible for transitioning the features of atom and bond, respectively. $X_{uv}^b$ signifies the bond feature between nodes $u$ and $v$, while $h_v^0 = X_v^a$ represents the original feature of node $v$. Details about the initial features of atoms and bonds are presented in Table 1. Specifically, the atom index for nodes at the molecule-level and motif-level are assigned values of 119 and 120, respectively, with their degrees set to zero. For both types of nodes, edges are augmented to represent atom-motif and motif-graph relationships. Notably, the bond feature indicating whether a bond is part of a ring 'is_in_ring' is not applicable in this context and is thus set to None. Missing features are treated as placeholder values during training. Our model is designed to learn robust representations even with incomplete data by leveraging the available features from other nodes and the graph structure. The model is able to infer the missing information through the learning process, especially with the help of neighborhood information and feature propagation mechanisms. Although the impact of missing features on downstream performance depends on the context, preliminary results show that HLN-DDI remains robust and competitive even in the case of such missing information.

Through the aforementioned encoding processes, we obtain the atomic representations $H_i^a$ and $H_j^a$, motif representations $H_i^m$ and $H_j^m$, as well as molecular representations $V_i^g$ and $V_j^g$ for the two drugs. Given that a drug typically comprises multiple atoms and motifs, we utilize the READOUT function to aggregate the features of the respective nodes corresponding to these atoms and motifs. In our model, the READOUT function is defined as a straightforward summation operation. Consequently, we have successfully derived atom-level representations $V_i^a$ and $V_j^a$, motif-level representations $V_i^m$ and $V_j^m$, and molecule-level representations $V_i^g$ and $V_j^g$ for the two drugs.

#### *Interaction importance measurement and prediction*

Once the representations of the two drugs at various levels have been obtained, co-attention mechanism was then applied to quantify the significance of each combination of these representations in predicting the final DDI. The scoring function within the co-attention mechanism is given by.

**Table 1** Atom and bond features

| Atom feature | Description |
|---|---|
| Atom type | [1,2,...,118]+[119,120] |
| Atomic degree | [0,1,2,3,4,5,6,7,8,9,10] |
| Bond feature | Description |
| Edge feature | {Single,double,triple,aromatic} |
| Is_in_ring | {False,True}+{None} |

$$r_{xy} = b^T tanh(W_x V_i^x + W_y V_j^y) \tag{3}$$

We use Xavier initialization [37] for these parameters: trainable weight vector $b$, weight matrices $W_x$, and weight matrices $W_y$. Xavier initialization is a widely adopted method in neural networks to maintain a stable gradient flow during training, which can help prevent gradient vanishing or explosion issues, leading to better convergence and improved training efficiency. For further reproducibility, we have cited standard implementations in popular deep learning frameworks such as PyTorch and TensorFlow. To independently modify the feature representation of drugs across various levels and bolster the representation of each individual drug, we introduce different weight matrices. This setup promotes the interaction between the two drugs by effectively leveraging the co-attention mechanism. Ultimately, the final prediction probability for the DDI within this triplet is derived based on the resultant co-attention map.

$$P(d_i, r, d_j) = \sigma\left(\sum_{V_i^x \epsilon V_i} \sum_{V_j^y \epsilon V_j} r_{xy} V_i^x M_r V_j^y\right) \tag{4}$$

In this context, $\sigma$ represents the sigmoid function, and $M_r$ is a trainable matrix that corresponds to the interaction type $r$. The sets $V_i = \{V_i^a, V_i^m, V_i^g\}$ and $V_j = \{V_j^a, V_j^m, V_j^g\}$ encapsulate the multi-level representations of drugs $d_i$ and $d_j$, respectively. Equation 4 views the probability of a drug pair $(d_i, d_j)$ exhibiting an interaction $r$ as a composite of paired chemical reactions across various representation levels of these drugs. Each of these paired reactions is assigned a weight reflecting its significance or relevance in predicting the final DDI.

The triplets present in the DDI dataset are regarded as positive samples. Corresponding to each positive sample, we construct a negative sample by randomly substituting either $d_i$ or $d_j$ within it, adhering to the methodology outlined by Wang et al. [38]. The performance of the model across the entire dataset $D$ is evaluated using binary cross-entropy loss [39], which is meticulously defined in Equation

$$L = -\frac{1}{N}\sum_{i=1}^{N}(log(p_i) + log(-p_i')) \tag{5}$$

Here, $N$ represents the sum of samples within the DDI dataset. Both $p_i$ and $p_i'$ are all computed by Eq. 4, the former corresponding to the predicted probability of a positive sample and the latter corresponding to the predicted probability of a negative sample.

## Experiments and results

### Datasets

**DrugBank dataset** [40]: DrugBank encompasss chemoinformatics, bioinformatics and other collections containing detailed drug data, comprises 1,704 drugs and 191,808 DDIs. There are 86 unique interaction types that detail the metabolic impact of one drug on another. There is one and only one interaction type for each pair of drugs in DrugBank.

**Twosides dataset**: Zitnik et al. [41] refined the original Twosides dataset [6] by removing side effects that could be attributed to individual drugs, retaining only statistically

significant adverse effects caused specifically by drug combinations. This filtering enhanced the reliability of the dataset for studying drug-drug interaction-induced side effects. Twosides focuses on phenotypic outcomes, and a drug pair in Twosides is likely associated with multiple interaction types. This dataset comprised 4,576,287 DDI tuples, 645 drugs and 963 interaction types between drug pairs.

### Experimental settings

We performed experiments under the transductive setting on the two datasets mentioned above and under the inductive setting only on the Drugbank dataset. In the context of transductive setting, each drug appearing in the DDI triplets of the test set is also included in the training set, guaranteeing that at least one interaction involving these drugs is already known. In contrast, under the inductive setting, the test set often includes DDI triplets involving one or two drugs with completely unknown interaction profiles with other drugs, which poses a greater challenge for predicting DDI compared to the transductive setting. It offers a more realistic representation of a model's applicability in predicting interactions for novel, unseen drugs. We train and evaluate the model based on the parameters presented in the Table 2. Our model is run for 300 epochs during the training phase under the transductive setting, where the learning rate for the first 200 epochs is 0.001, and the learning rate for the last 100 epochs is 0.0001. In the inductive setting, the number of training epoch is 100, all with the learning rate of 0.0001. The activation function of MLP uses relu. The following four metrics are adopted to evaluate the performance of the models: Accuracy (ACC), Area Under the ROC Curve (AUC), F1 value (F1), and average accuracy (AP).

### Performance of HLN-DDI in comparative experiments

We evaluated the performance of HLN-DDI against several state-of-the-art methodologies sharing similarities with our proposed approach. Specifically, all comparison methods: (1) utilize molecular graphs as primary inputs, (2) maintain a consistent benchmark DDI dataset structure, (3) consider substructure influences during drug representation learning, and (4) apply to both transductive and inductive learning scenarios. The methods selected for comparison are as follows:

- **GMPNN-CS** [18]: Extracts substructures of varying sizes and shapes from molecular graphs and predicts DDIs based on interactions between paired substructures.

**Table 2** Parameters of model experiments

| Parameters | Values | |
|---|---|---|
| | Transductive | Inductive |
| Node embedding dimension | 64 | 64 |
| Number of Graph Isomorphism Network layers | 5 | 5 |
| Dropout rate of Graph Isomorphism Network | 0.2 | 0.4 |
| Epoch | 300 | 100 |
| Batch size | 256 | 256 |

- **DGNN-DDI** [19]: Utilizes a directed graph neural network for message passing, incorporates substructure attention mechanisms, and employs co-attention to weight interaction scores among substructures.
- **MSAN** [23]: Uses a Transformer-like module to generate fixed representative vectors for substructure patterns, with predictions based on substructure interaction strengths combined with drug features.
- **SSI-DDI** [20]: Considers each node's hidden feature as a substructure and computes interactions among these features for final DDI predictions.
- **SA-DDI** [17]: Employs a message-passing neural network with a dedicated substructure-substructure interaction module to derive informative drug features.
- **DSN-DDI** [21]: Iteratively captures drug substructure knowledge through intra-view and inter-view perspectives and predicts DDIs based on substructure interactions.

#### *Performance evaluation under the transductive setting*

Following the experimental protocol proposed by Nyamabo et al. [18], all triples in our dataset were partitioned into training, validation, and testing sets using a 6:2:2 ratio. To ensure robustness, this partitioning strategy was repeated three times to generate three stratified random folds. All comparative methods utilized identical training, validation, and testing sets for fair evaluation.

Tables 3 and 4 summarize the comparative performance results, demonstrating that our proposed model, HLN-DDI, consistently outperforms existing state-of-the-art methods across multiple evaluation metrics. Specifically, HLN-DDI achieved performance scores surpassing 99% on most metrics when evaluated on both the DrugBank dataset and the larger, more complex Twosides dataset, underscoring its exceptional capability in predicting DDIs. Compared to other robust models previously established for their high accuracy within transductive settings, HLN-DDI demonstrates a notable improvement of 1.21% in accuracy (ACC) over the second-best performing method. These results highlight the strong predictive accuracy and reliability of HLN-DDI, validating its effectiveness in drug-drug interaction prediction tasks.

**Table 3** Performance assessment of HLN-DDI and comparison baselines under transductive setting on DrugBank dataset

| Method | ACC | AUC | AP | F1 |
|---|---|---|---|---|
| GMPNN-CS | 95.30 ± 0.05 | 98.46 ± 0.01 | 97.94 ± 0.02 | 95.39 ± 0.05 |
| DGNN-DDI | 96.09 ± 0.27 | 98.94 ± 0.31 | 98.51 ± 0.37 | 96.16 ± 0.31 |
| MSAN | 96.69 ± 0.61 | 99.27 ± 0.78 | 99.12 ± 1.00 | 97.04 ± 0.60 |
| SSI-DDI | 96.33 ± 0.09 | 98.95 ± 0.08 | 98.57 ± 0.14 | 96.38 ± 0.09 |
| SA-DDI | 96.23 ± 0.01 | 98.80 ± 0.02 | 98.36 ± 0.04 | 96.29 ± 0.09 |
| DSN-DDI | 96.94 ± 0.02 | 99.47 ± 0.01 | 99.37 ± 0.02 | 96.93 ± 0.02 |
| w/o_atom | 97.14 ± 0.01 | 99.17 ± 0.12 | 99.27 ± 0.28 | 96.13 ± 0.41 |
| w/o_motif | 96.83 ± 0.11 | 98.79 ± 0.08 | 98.43 ± 0.50 | 97.13 ± 0.17 |
| w/o_molecule | 97.55 ± 0.08 | 98.94 ± 0.13 | 98.61 ± 0.39 | 96.48 ± 0.45 |
| HLN-DDI | 98.15 ± 0.01 | 99.67 ± 0.46 | 99.86 ± 0.03 | 97.64 ± 0.04 |

**Table 4** Performance assessment of HLN-DDI and comparison baselines under transductive setting on Twosides dataset

| Method | ACC | AUC | AP | F1 |
|---|---|---|---|---|
| GMPNN-CS | 82.83 ± 0.14 | 90.07 ± 0.12 | 87.24 ± 0.12 | 84.08 ± 0.14 |
| DGNN-DDI | 85.29 ± 0.94 | 91.92 ± 1.08 | 89.41 ± 1.45 | 86.12 ± 0.65 |
| MSAN | 80.45 ± 1.41 | 88.52 ± 1.63 | 84.32 ± 1.95 | 77.88 ± 0.85 |
| SSI-DDI | 78.20 ± 0.14 | 85.85 ± 0.13 | 82.71 ± 0.14 | 79.81 ± 0.16 |
| SA-DDI | 87.45 ± 0.03 | 93.17 ± 0.04 | 90.51 ± 0.08 | 88.35 ± 0.04 |
| DSN-DDI | 98.83 ± 0.04 | 99.90 ± 0.01 | 99.89 ± 0.01 | 98.83 ± 0.04 |
| w/o_atom | 98.29 ± 0.23 | 98.71 ± 0.26 | 98.89 ± 0.33 | 98.76 ± 0.31 |
| w/o_motif | 98.13 ± 0.56 | 99.56 ± 0.03 | 99.13 ± 0.33 | 99.00 ± 0.01 |
| w/o_molecule | 98.40 ± 0.31 | 99.30 ± 0.20 | 99.47 ± 0.56 | 98.83 ± 0.05 |
| HLN-DDI | 99.13 ± 0.09 | 99.93 ± 0.05 | 99.91 ± 0.11 | 99.04 ± 0.02 |

**Table 5** Performance assessment of HLN-DDI and comparison baselines on DrugBank dataset under the P1 partition in the inductive setting

| Method | ACC | AUC | AP | F1 |
|---|---|---|---|---|
| GMPNN-CS | 68.57 ± 0.30 | 74.96 ± 0.40 | 75.44 ± 0.50 | 65.32 ± 0.23 |
| DGNN-DDI | 70.31 ± 0.62 | 72.87 ± 0.92 | 73.65 ± 1.61 | 67.21 ± 0.59 |
| MSAN | 66.31 ± 0.61 | 72.75 ± 0.78 | 71.61 ± 1.00 | 68.68 ± 0.60 |
| SSI-DDI | 65.40 ± 1.30 | 73.43 ± 1.81 | 75.03 ± 1.42 | 54.12 ± 3.46 |
| SA-DDI | 67.15 ± 0.08 | 73.62 ± 1.25 | 73.39 ± 1.40 | 63.40 ± 1.53 |
| DSN-DDI | 73.42 ± 1.29 | 81.79 ± 1.12 | 81.82 ± 1.48 | 70.34 ± 0.98 |
| w/o_atom | 73.12 ± 0.61 | 82.91 ± 1.33 | 82.27 ± 0.73 | 72.83 ± 0.61 |
| w/o_motif | 73.89 ± 0.41 | 81.99 ± 0.48 | 83.08 ± 1.35 | 71.99 ± 0.75 |
| w/o_molecule | 74.59 ± 0.22 | 82.70 ± 0.76 | 83.59 ± 1.39 | 73.04 ± 1.00 |
| HLN-DDI | 75.44 ± 0.11 | 83.25 ± 0.27 | 83.79 ± 0.96 | 73.83 ± 0.46 |

#### *Performance evaluation under the inductive setting*

Under the inductive setting, 20% of the drugs in the DrugBank dataset were randomly selected as new (unknown) drugs, and all interaction data involving these drugs were excluded from the training process. The remaining drugs were considered existing (known) drugs. For inductive evaluation, two distinct partitioning strategies were employed: partition scheme P1, where each DDI triple in the test set consists exclusively of pairs of new drugs, and partition scheme P2, where each DDI triple in the test set involves one existing drug and one new drug. The remaining data, excluding the test set, were used as the training set.

To ensure robustness, we conducted three independent experiments using 3-fold cross-validation (CV) on the DrugBank dataset under both inductive scenarios. The results of these evaluations are summarized in Tables 5 and 6. HLN-DDI consistently delivered superior performance across all evaluation metrics in both partitioning schemes, demonstrating its strong capability in addressing the challenge posed by limited interaction data for novel drugs. Additionally, HLN-DDI significantly outperformed the other comparison methods, reinforcing its effectiveness and adaptability in inductive drug-drug interaction prediction.

**Table 6** Performance assessment of HLN-DDI and comparison baselines on DrugBank dataset under the P2 partition in the inductive setting

| Method | ACC | AUC | AP | F1 |
|---|---|---|---|---|
| GMPNN-CS | 77.72 ± 0.30 | 84.84 ± 0.15 | 84.87 ± 0.40 | 78.29 ± 0.16 |
| DGNN-DDI | 77.07 ± 0.94 | 80.35 ± 1.08 | 82.97 ± 1.45 | 73.03 ± 0.65 |
| MSAN | 69.83 ± 1.41 | 77.29 ± 1.63 | 75.79 ± 1.95 | 73.01 ± 0.85 |
| SSI-DDI | 76.38 ± 0.92 | 84.23 ± 1.05 | 84.94 ± 0.76 | 73.54 ± 1.50 |
| SA-DDI | 75.55 ± 1.12 | 82.95 ± 1.05 | 84.11 ± 0.92 | 71.94 ± 1.57 |
| DSN-DDI | 81.92 ± 1.20 | 91.01 ± 0.76 | 91.09 ± 0.93 | 80.18 ± 1.49 |
| w/o_atom | 82.49 ± 0.73 | 90.93 ± 1.12 | 92.00 ± 0.98 | 80.86 ± 0.23 |
| w/o_motif | 82.90 ± 0.69 | 91.79 ± 0.88 | 91.84 ± 1.50 | 81.36 ± 0.61 |
| w/o_molecule | 83.00 ± 1.11 | 91.12 ± 1.06 | 91.57 ± 1.15 | 81.29 ± 0.40 |
| HLN-DDI | 83.38 ± 0.41 | 91.98 ± 0.83 | 92.20 ± 0.74 | 81.97 ± 0.27 |

**Table 7** Model setups in ablation experiments, with a check mark indicating that the representation under that structural level participates in the final DDI prediction, otherwise it is removed

| Ablation model | Atom-level | Motif-level | Molecule-level |
|---|---|---|---|
| wo_atom | | ✓ | ✓ |
| wo_motif | ✓ | | ✓ |
| wo_molecule | ✓ | ✓ | |

### Ablation experiment

To explore the necessity of HLN-DDI in learning drug representations at each hierarchical level and to verify the benefits of multi-level representation fusion, we conducted an ablation experiment. In each experiment, we reconfigured three different comparative models by removing the drug representations at a specified level while retaining the representations at the other levels, as detailed in Table 7.

We subsequently reported the experimental results of the 3-fold cross-validation (CV) conducted under transductive and inductive settings on the DrugBank dataset, as well as under the transductive setting on the Twosides dataset. As illustrated in Tables 3, 4, 5, and 6, each comparative model involved in the ablation experiments demonstrated competent performance in the DDI prediction tasks, indicating the relevance and informativeness of the constructed hierarchical features. Moreover, the complete HLN-DDI framework consistently outperformed all its variants, underscoring the advantage of integrating drug information from all three structural levels to derive more comprehensive drug representations.

### Drug feature visualization

To further explore and interpret the hierarchical drug features extracted by HLN-DDI, we visualized and compared the encoded feature representations of drug pairs across different structural levels. Specifically, we concatenated the feature vectors of drug pairs at corresponding hierarchical levels and applied the t-distributed Stochastic Neighbor Embedding (t-SNE) [42] method for dimensionality reduction and visualization. We then compared these processed features against their original, unencoded counterparts.

The visualization results comparing original drug pairings with features encoded by HLN-DDI at three structural levels-atom-level (Fig. 3A,D), motif-level (Fig. 3B,E), and molecular-level (Fig. 3C,F)-are presented in Fig. 3. In these visualizations, red points indicate drug pairs known to interact (positive samples), while blue points represent pairs without known interactions (negative samples). Notably, after feature encoding through HLN-DDI, the positive and negative samples distinctly separate, forming clear and identifiable clusters at each hierarchical level.

These visualization results highlight that drug-pair feature representations become significantly more discriminative after processing by our proposed framework compared to their original states. This clear separation demonstrates that the integration of hierarchical feature learning and a co-attention mechanism within HLN-DDI substantially improves the discriminability of drug-pair interactions, further validating the importance of hierarchical representation enriched by co-attention mechanisms for accurate DDI prediction.

### Real-world DDI applications

To evaluate the practical applicability of HLN-DDI, we conducted experiments using the DrugBank dataset. Specifically, we trained our model on data involving previously approved (old) drugs to assess its predictive performance on interactions involving newly FDA-approved drugs. Using FDA-approval data [43], we classified all drugs within the DrugBank dataset into two categories based on whether their FDA approval occurred before or after 2017. Table 8 summarizes the data partitioning. The training set included DDI triplets exclusively involving drugs approved before 2017, whereas the test set consisted of DDI triplets with at least one drug approved after 2017.

To provide a more comprehensive assessment of the performance of our HLN-DDI model, we have incorporated additional competitive baselines in our experiments. In

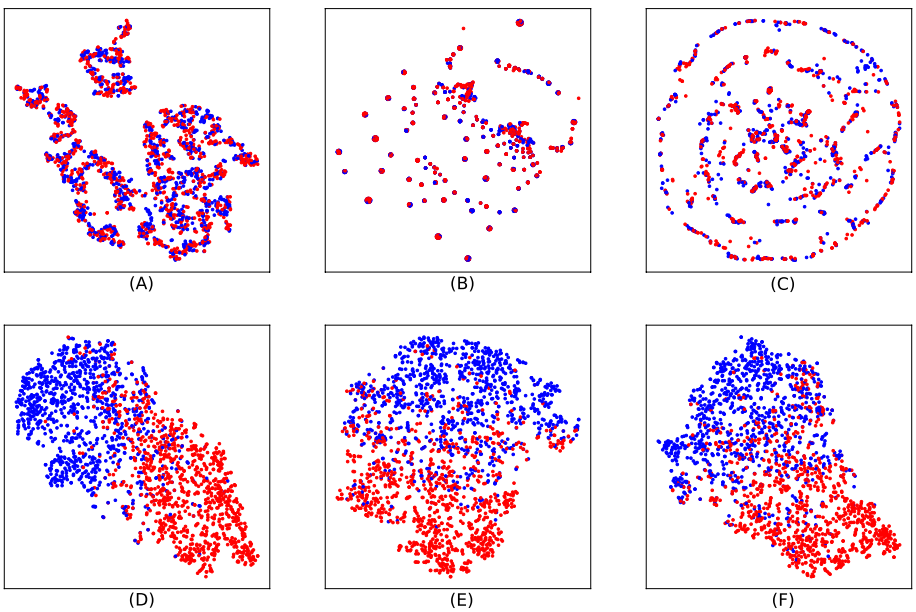


**Fig. 3** Visualization results of drug pair features after t-SNE processing at atom-level (**A**,**D**), motif-level (**B**,**E**), and molecule-level (**C**,**F**)

**Table 8** The statistics of the DrugBank dataset for new drug application

| Drug split | Drugs | DDIs |
|---|---|---|
| Old drugs before 2017 | 1681 | 188341 |
| New drugs after 2017 | 25 | 3467 |
| Total | 1706 | 191808 |

addition to the previously used models, we now compare HLN-DDI with a wider range of methods, including traditional machine learning approaches such as Random Forests and Support Vector Machines (SVM), and as detailed in Table 9. Our results demonstrate that HLN-DDI outperforms these baseline models in several key metrics, particularly in capturing complex, hierarchical drug interaction patterns, and providing robust predictions across different drug categories. Additionally, the area under the ROC curves shown in Fig. 4 further illustrates HLN-DDI's enhanced capability to differentiate positive and negative drug interaction pairs effectively.

In real-world healthcare settings, the deployment of HLN-DDI faces several challenges, particularly regarding computational resource requirements and inference

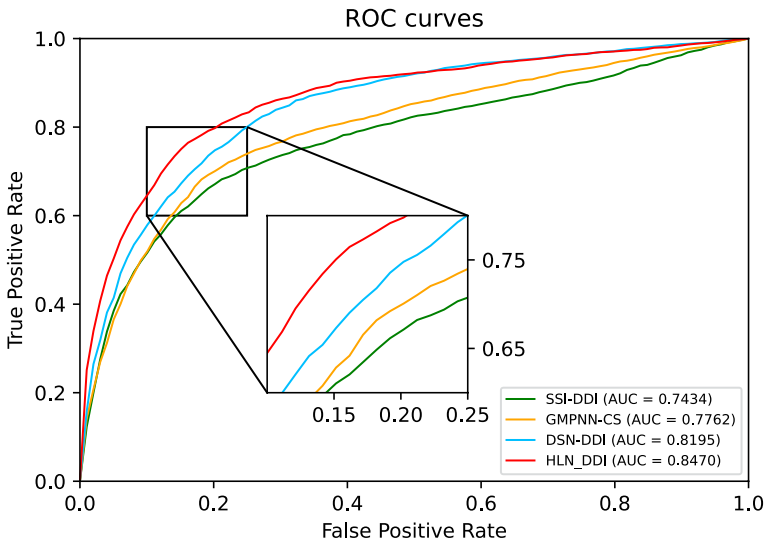


**Fig. 4** Performance evaluation of SSI-DDI, GMPNN-CS, DSN-DDI and HLN-DDI for new FDA-approved drugs

**Table 9** Performance evaluation between SSI-DDI, GMPNN-CS, DSN-DDI and HLN-DDI on DrugBank dataset for new drug application

| Method | ACC | AUC | AP | F1 |
|---|---|---|---|---|
| SVM | 58.32 | 66.92 | 67.03 | 63.84 |
| Random forest | 60.71 | 70.35 | 71.26 | 64.99 |
| DGNN-DDI | 68.02 | 75.59 | 74.46 | 67.15 |
| MSAN | 67.74 | 73.99 | 72.17 | 64.83 |
| SA-DDI | 67.44 | 76.23 | 73.00 | 63.45 |
| SSI-DDI | 67.85 | 74.34 | 73.18 | 68.32 |
| GMPNN-CS | 68.10 | 77.62 | 73.64 | 64.25 |
| DSN-DDI | 74.07 | 81.95 | 78.95 | 77.22 |
| HLN-DDI | 76.09 | 84.70 | 80.01 | 79.42 |

latency. Due to the complexity of the deep learning architecture, inference speed may become a bottleneck in clinical applications. Therefore, reducing computational demands and inference latency is crucial for broader adoption. Future work could explore model lightweighting techniques, such as pruning and quantization, to lower the computational overhead. Additionally, leveraging hardware acceleration and distributed computing could further enhance real-time inference capabilities, making HLN-DDI more suitable for clinical environments.

## Discussion

Drug-drug interactions (DDIs) play a pivotal role in pharmacology, significantly influencing therapeutic outcomes by either enhancing beneficial effects or causing adverse reactions when multiple medications are co-administered. Accurate prediction of DDIs is essential to ensure the safety and efficacy of combined drug therapies. Given the clinical significance of this task, numerous in vitro and computational methods have been proposed as viable alternatives for predicting DDIs. The key innovations and advantages of our HLN-DDI model compared to existing methods are summarized below:

(1) **Enhanced substructure decomposition**: The molecular decomposition strategy implemented in HLN-DDI achieves a more precise and detailed structural breakdown while preserving the chemical significance and interpretability of the resulting substructures.
(2) **Effective molecular graph augmentation**: Unlike other augmentation methods that risk disrupting molecular structural integrity, the augmentation approach in HLN-DDI carefully preserves essential molecular characteristics, resulting in reliable and informative feature extraction.
(3) **Hierarchical node interaction**: HLN-DDI incorporates nodes from atom-level, motif-level, and molecule-level hierarchies during the encoding phase, allowing bidirectional message passing. This strategy effectively overcomes the conventional limitation of unidirectional information flow, resulting in more comprehensive molecular representations.
(4) **Explainable and comprehensive integration**: Our model systematically integrates hierarchical information from both drugs using an interpretable co-attention mechanism. This comprehensive approach ensures that the final predictions are rational and derived from a thorough set of chemically meaningful features, enhancing the reliability and interpretability of DDI predictions.

While the model demonstrates strong generalizability, its performance may vary across different classes of drugs or specific interaction types. For example, drug classes with limited training data or highly complex interactions might lead to reduced prediction accuracy. This could indicate potential domain-specific biases or limitations in the model's ability to capture all possible drug interactions. To address these issues, future work can explore data enhancement techniques or incorporate additional domain knowledge to improve the performance of models on underrepresented drug

classes or interaction types. We can also study how to enhance the scalability of the model and integrate other data sources, such as pharmacological and clinical datasets, to expand the versatility of hln-ddi and improve its practical applicability.

## Conclusion

In this study, we introduced a novel hierarchical learning network, HLN-DDI, which leverages molecular structure representations and a co-attention mechanism for drug-drug interaction prediction. HLN-DDI simultaneously captures atom-level, motif-level, and molecule-level features from each drug, subsequently integrating these multi-level representations through a co-attention mechanism. This integration assesses the importance of representation combinations to calculate the probability of potential DDIs.

Extensive experimental evaluations on DrugBank and Twosides datasets consistently demonstrated the superior performance of HLN-DDI relative to state-of-the-art methods. Furthermore, ablation studies highlighted the critical importance of extracting drug features across multiple hierarchical levels. Visualization analyses provided intuitive confirmation that the extracted features at each hierarchical level are both meaningful and highly distinguishable. The successful application of HLN-DDI to real-world DDI scenarios further underscores the practical relevance and utility of our model.

In conclusion, HLN-DDI has the potential to significantly accelerate drug discovery by improving the efficiency and accuracy of drug-target interaction predictions. By addressing key clinical demands, such as the need for faster and more precise drug conjugation evaluations, this method can streamline the drug development process and reduce time-to-market. The alignment of HLN-DDI with these clinical needs underscores its potential to contribute not only to scientific advancement but also to improving patient outcomes and healthcare delivery on a global scale. Thus, the model holds great promise in shaping the future of personalized medicine and drug development.

**Abbreviations**

| | |
|---|---|
| DDI | Drug-drug interactions |
| ADR | Adverse drug reaction |
| BRISC | Breakdown of recurrent chemical substructures |
| GNN | Graph neural network |
| GAT | Graph attention networks |
| GCN | Graph convolutional network |
| GIN | Graph isomorphism network |
| MLP | Multilayer perceptro |
| AUC | The area under the ROC curve |
| ACC | Accuracy |
| AP | Average precision |
| CV | Cross-validation |
| FDA | Food and drug administration |

**Acknowledgements**

We are grateful for resources from the High Performance Computing Center of Central South University.

**Author contributions**

Y.L. conceptualized the main ideas and framed the manuscript. L.D. evaluated and supervised the conceptualization of the article. Y.L. designed the experiments, constructed the experimental environment, conducted the experiments, collected the data, and analyzed the results. L.D. and Z.H revised the manuscript and L.D. provided funding. All authors wrote, revised and approved the manuscript.

**Funding**

This work was supported by the National Natural Science Foundation of China under grants No. U23A20321 and No.62272490.

**Availability of data and materials**
The code and dataset for this study can be found at https://github.com/lylikeeMoon/HLN-DDI.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no conflict of interest.

Received: 21 March 2025 Accepted: 6 May 2025
Published online: 04 June 2025

**References**

1. Dale MM, Haylett DG. Rang & Dale's pharmacology flash cards updated edition e-book. Philadelphia, USA: Elsevier Health Sciences; 2013.
2. Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. Nat Rev Drug Discov. 2005;4(10):825–33.
3. Juurlink DN, Mamdani M, Kopp A, Laupacis A, Redelmeier DA. Drug-drug interactions among elderly patients hospitalized for drug toxicity. JAMA. 2003;289(13):1652–8.
4. Yeh P, Tschumi AI, Kishony R. Functional classification of drugs by properties of their pairwise interactions. Nat Genet. 2006;38(4):489–94.
5. Ryall KA, Tan AC. Systems biology approaches for advancing the discovery of effective drug combinations. J Cheminformatics. 2015;7:1–15.
6. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012;4(125):125–3112531.
7. Hong E, Jeon J, Kim HU. Recent development of machine learning models for the prediction of drug-drug interactions. Korean J Chem Eng. 2023;40(2):276–85.
8. Zhao X-M, Iskar M, Zeller G, Kuhn M, Van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. PLoS Comput Biol. 2011;7(12):1002323.
9. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform. 2016;17(1):2–12.
10. Rudrapal M, Khairnar SJ, Jadhav AG. Drug repurposing (dr): an emerging approach in drug discovery. In: Badria, F.A. (ed.) Drug Repurposing. IntechOpen, Rijeka 2020. Chap. 1. https://doi.org/10.5772/intechopen.93193 .
11. Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, Xiao G, Li Y, Allen J, Zhong R, et al. A community computational challenge to predict the activity of pairs of compounds. Nat Biotechnol. 2014;32(12):1213–22.
12. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang G-Z. Deep learning for health informatics. IEEE J Biomed Health Inf. 2016;21(1):4–21.
13. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M. Graph neural networks: a review of methods and applications. AI Open. 2020;1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001.
14. Xu N, Wang P, Chen L, Tao J, Zhao J. Mr-gnn: multi-resolution and dual graph neural network for predicting structured entity interactions, 2019. arXiv preprint arXiv:1905.09558.
15. Zhang X, Wang G, Meng X, Wang S, Zhang Y, Rodriguez-Paton A, Wang J, Wang X. Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug-drug interactions prediction. Brief Bioinform. 2022;23(5):296.
16. Deac A, Huang Y-H, Veličković P, Liò P, Tang J. Drug-drug adverse effect prediction with graph co-attention, 2019. arXiv preprint arXiv:1905.00534.
17. 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.
18. Nyamabo AK, Yu H, Liu Z, Shi J-Y. Drug-drug interaction prediction with learnable size-adaptive molecular substructures. Brief Bioinform. 2022;23(1):441.
19. 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):1010812.
20. Nyamabo AK, Yu H, Shi J-Y. SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction. Brief Bioinform. 2021;22(6):133.
21. Li Z, Zhu S, Shao B, Zeng X, Wang T, Liu T-Y. DSN-DDI: an accurate and generalized framework for drug-drug interaction prediction by dual-view representation learning. Brief Bioinform. 2023;24(1):597.
22. Harrold MW, Zavod RM. Basic concepts in medicinal chemistry. Abingdon: Taylor & Francis; 2014.
23. Zhu X, Shen Y, Lu W. Molecular substructure-aware network for drug-drug interaction prediction. In: Proceedings of the 31st ACM international conference on information & knowledge management, pp. 2022; 4757–4761.
24. Duval A, Malliaros F. Higher-order clustering and pooling for graph neural networks. In: Proceedings of the 31st ACM international conference on information & knowledge management, pp. 2022;426–435.
25. Yu Z, Gao H. Molecular representation learning via heterogeneous motif graph neural networks. In: International conference on machine learning, pp. 2022;25581–25594. PMLR.

26. Degen J, Wegscheid-Gerlach C, Zaliani A, Rarey M. On the art of compiling and using'drug-like'chemical fragment spaces. ChemMedChem. 2008;3(10):1503.
27. Lu J, Yang J, Batra D, Parikh D. Hierarchical question-image co-attention for visual question answering. Advances in neural information processing systems. 2016;29.
28. 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.
29. Landrum G, et al. Rdkit: a software suite for cheminformatics, computational chemistry, and predictive modeling. Greg Landrum. 2013;8(31.10):5281.
30. Zang X, Zhao X, Tang B. Hierarchical molecular graph self-supervised learning for property prediction. Commun Chem. 2023;6(1):34.
31. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: International conference on machine learning, pp. 2017;1263–1272. PMLR.
32. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: International conference on learning representations 2016.
33. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, et al. Graph attention networks. Statistics. 2017;1050(20):10–48550.
34. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? In: International conference on learning representations 2018.
35. Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. Graphdta: predicting drug-target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140–7.
36. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2020;32(1):4–24.
37. 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, JMLR Workshop and Conference Proceedings pp. 2010;249–256.
38. Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI conference on artificial intelligence, vol. 2014;28.
39. Mao A, Mohri M, Zhong Y. Cross-entropy loss functions: theoretical analysis and applications. In: International conference on machine learning, pp. 2023;23803–23828. PMLR.
40. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 2018;46(D1):1074–82.
41. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):457–66.
42. Maaten L, Hinton G. Visualizing data using t-sne. Journal of machine learning research. 2008;9(11).
43. Mullard A. FDA drug approvals. Nat Rev Drug Discov. 2020. https://doi.org/10.1038/d41573-021-00002-0.


## Publisher's Note