

RESEARCH

Open Access



HDN-DDI: a novel framework for predicting drug-drug interactions using hierarchical molecular graphs and enhanced dual-view representation learning

Jinchen Sun^{1,2} and Haoran Zheng^{1,2*}

*Correspondence:
hrzheng@ustc.edu.cn

¹ School of Computer Science and Technology, University of Science and Technology of China, 443 Huangshan Road, Hefei 230027, China

² Anhui Key Laboratory of Software Engineering in Computing and Communication, University of Science and Technology of China, 443 Huangshan Road, Hefei 230027, China

Abstract

Background: Drug–drug interactions (DDIs) especially antagonistic ones present significant risks to patient safety, underscoring the urgent need for reliable prediction methods. Recently, substructure-based DDI prediction has garnered much attention due to the dominant influence of functional groups and substructures on drug properties. However, existing approaches face challenges regarding the insufficient interpretability of identified substructures and the isolation of chemical substructures.

Results: This study introduces a novel framework for DDI prediction termed HDN-DDI. HDN-DDI integrates an explainable substructure extraction module to decompose drug molecules and represents them using innovative hierarchical molecular graphs, which effectively incorporates information from real chemical substructures and improves molecules encoding efficiency. Furthermore, the enhanced dual-view learning method inspired by the underlying mechanisms of DDIs enables HDN-DDI to comprehensively capture both hierarchical structure and interaction information. Experimental results demonstrate that HDN-DDI has achieved state-of-the-art performance with accuracies of 97.90% and 99.38% on the two widely-used datasets in the warm-start setting. Moreover, HDN-DDI exhibits substantial improvements in the cold-start setting with boosts of 4.96% in accuracy and 7.08% in F1 score on previously unseen drugs. Real-world applications further highlight HDN-DDI's robust generalization capabilities towards newly approved drugs.

Conclusion: With its accurate predictions and robust generalization across different settings, HDN-DDI shows promise for enhancing drug safety and efficacy. Future research will focus on refining decomposition rules as well as integrating external knowledge while preserving the model's generalization capabilities.

Keywords: Drug-drug interactions, Chemical substructures, Hierarchical molecular graphs, Graph attention network, Dual-view representation learning



Introduction

In the treatment of complex diseases, the co-administration of two or more drugs can target various biological processes implicated in disease development thereby achieving therapeutic effects that are unattainable with monotherapy [1–3]. However, drug-drug interactions (DDIs) such as antagonism between different drugs potentially lead to severe side effects or adverse drug events [4–6]. Therefore, it is crucial to identify these potential DDIs.

Traditional methods which rely on pharmaceutical research and clinical trials are both costly and time-consuming when applied to numerous potential drug combinations [7–9]. Computational-based methods offer a more cost-effective and expedited means of identification [10–12]. In recent years, machine learning and deep learning-based methods have exhibited superior performance in predicting DDIs [13–15]. Molecular graph-based methods have especially attracted interest for their ability to adapt to new drugs without relying on external knowledge [16–18]. These methods represent drug molecules using undirected graphs where atoms and chemical bonds are represented as nodes and edges respectively. Subsequently, a variety of graph neural networks (GNNs) are utilized for molecular representation learning and DDI prediction [19–21].

Given that drugs comprise various functional groups or substructures that determine their pharmacodynamics and pharmacokinetic properties and ultimately influence their interactions [22, 23], substructure-based methods have been successively proposed [24–27]. Yang et al. [24] introduced a substructure-aware network SA-DDI to capture size-adaptive and shape-adaptive substructures. Nyamabo et al. [25] explicitly decomposed DDI prediction into substructure-substructure interactions prediction and employed a gated message passing network to learn substructures of different sizes and shapes. Li et al. [26] proposed a dual-view representation learning framework DSN-DDI to learn substructures through intra-molecular and inter-molecular views. Ning et al. [27] further developed the framework with a bilinear representation extraction layer. However, there exist differences between the substructures identified by models and those known in medicinal chemistry and the differences require supplementary manual interpretation. In essence, current methods suffer from insufficient interpretability of identified substructures.

In addition to data-driven methods, molecule decomposition based on chemical rules has been gradually attracting attention [28, 29]. For instance, Zang et al. [29] presented HiMol which employed the BRICS algorithm [30] and additional chemical rules to decompose molecules. Furthermore, it learned molecular representation using novel hierarchical molecular graphs and achieved superior performance in molecular property prediction. Fan et al. [31] proposed PMSE-DDI that utilizes the BRICS algorithm to fragment drug molecules into chemical substructures and encodes them to predict DDIs. Notably, the substructures employed in PMSE-DDI adhere to chemical rules without the need for additional manual explanation. However, chemical substructures within the same molecule are isolated from each other, which prevents them from accessing both local and global information about the molecule.

To address the limitations of existing methods such as the insufficient interpretability of identified substructures and the isolation of chemical substructures, we introduced a novel framework named HDN-DDI. This framework utilizes hierarchical molecular

graphs and enhanced dual-view representation learning method for DDI prediction. The main innovations of this framework are as follows:

- An explainable substructure extraction module that decomposes drug molecules using an interpretable chemical decomposition algorithm, which fundamentally enhances the interpretability of substructures.
- A novel way for molecular representation learning, which fully integrates hierarchical structure and the information from real chemical substructures in the molecules.
- An enhanced dual-view representation learning method that focuses on key substructures, which aligns more with the actual process of DDIs and effectively improve the model's performance.

The experimental results on two widely-used datasets demonstrate that HDN-DDI outperforms existing models in both warm-start and cold-start settings. In the warm-start setting, HDN-DDI achieves an accuracy of 97.93% and an AUPR of 99.72% on the Drug-Bank dataset along with an accuracy of 99.38% and an AUPR of 99.97% on the Twosides dataset. In the cold-start setting, HDN-DDI achieves accuracies of 79.84% and 89.43% on two partitions, showcasing significant improvements of 4.96% and 4.08% respectively. Moreover, HDN-DDI also has significant improvements in other metrics including AUROC, AUPR and F1 score. These results collectively indicate the ability of HDN-DDI to generate more precise DDI predictions. Additionally, HDN-DDI's recognition accuracy surpasses that of the current state-of-the-art by 5.79% in real-world applications, further validating its effectiveness in predicting DDIs involving novel drugs.

Method

This section begins by formally defining the problem to be solved. Subsequently, it introduces each module within the framework including the input, the substructure extraction module, the HDN Encoder and the HDN Decoder. Finally, it illustrates the loss function employed in the framework.

Problem definition

Given a set of drugs $\mathcal{D}$ and a set of interaction types $\mathcal{R}$, the DDI prediction can be formalized as a function: $f : \mathcal{D} \times \mathcal{R} \times \mathcal{D} \rightarrow [0, 1]$. Herein, 1 means that the tuple (D_x, r, D_y) is valid which signifies the presence of an interaction r between the drugs D_x and D_y while 0 indicates the absence of such a tuple. The primary objective of this paper is to find an approximation function for f and predict the existence of the corresponding tuple given two drugs and an interaction type.

Input

The chemical structure of a drug molecule is initially represented by an undirected graph $G = (V, E)$ where V is the set of all nodes representing atoms in the molecule and $E \subset V \times V$ is the set of all edges representing the chemical bonds between atoms. Each node in the graph has an original 66-dimensional vector $x \in \mathbb{R}^{66}$ for input and the detailed chemical meanings of nodes are provided in Additional file 1. Consistent with

previous studies[23, 24, 26], edges in the graph simply represent neighboring relationships between nodes, enabling information to flow between them.

Overview of HDN-DDI

HDN-DDI initially decomposes drug molecules using an explainable substructure extraction module and then obtains representations of drugs at varying depths through the HDN Encoder. Finally, it aggregates the layer-wise representations of drugs and predicts DDIs using the HDN Decoder. The overall framework of HDN-DDI is illustrated in Fig. 1A.

Specifically, given the molecular graphs of drugs D_x and D_y , HDN-DDI decomposes them into several chemical substructures using the refined BRICS algorithm [29] and transforms the original molecular graphs into atomic-substructure-molecular hierarchical molecular graphs G_x and G_y . The HDN Encoder composed of several HDN Blocks accepts the hierarchical molecular graphs of both drugs as input. As depicted in Fig. 1B, each HDN Block learns the hierarchical representation from respective molecules and the interaction representation from both molecules and updates the hierarchical molecular graphs using the Graph Attention Network (GAT) [32]. The updated representation of molecular-level nodes is taken as the global molecular representations. The HDN

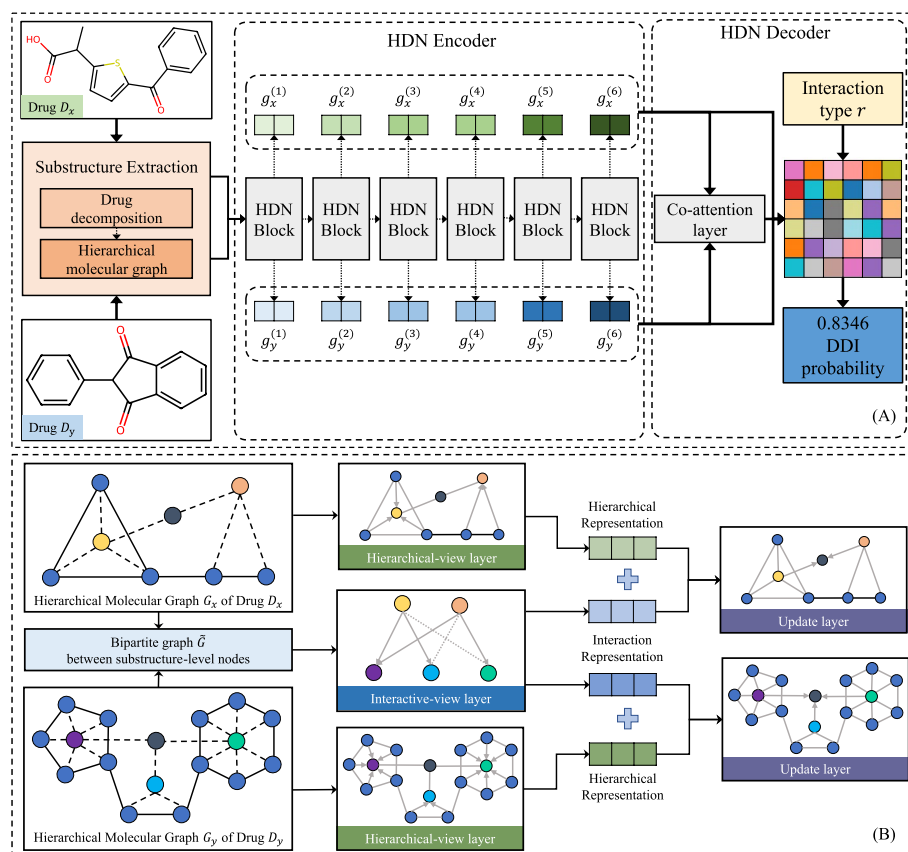


Fig. 1 The general sketch of HDN-DDI. **A** The overall framework of HDN-DDI. **B** The learning process of the HDN Block

Decoder employs a co-attention mechanism to integrate drug representations of different depths and makes the final DDI prediction.

Substructure extraction module

In the substructure extraction module, we utilize the refined BRICS algorithm[29] to decompose drug molecules and represent them using novel hierarchical molecular graphs. The processing flow of the module is illustrated in Fig. 2. Given a drug's SMILES (Simplified Molecular Input Line Entry System) sequence, we convert it into a molecular graph using RDKit [33] and apply the BRICS algorithm [30] to preliminarily decompose the graph into fragments. Any large ring fragments that cannot be decomposed by the BRICS algorithm are further partitioned into several smaller ring fragments based on additional rules [29]. As a result, we decompose an entire drug molecule into several chemical substructures that reveal the drug's properties as depicted in Fig. 2A.

Following the method proposed by Zang et al. [29], we construct the hierarchical molecular graph from the original graph based on the decomposition results, as illustrated in Fig. 2B. Initially, for each chemical substructure identified in the decomposition, we introduce a corresponding substructure-level node into the original graph which initially contains only atomic-level nodes. We then establish bidirectional edges between the substructure-level nodes and the atomic-level nodes based on the inclusion relationships between the corresponding substructures and their constituent atoms. For example, if a substructure S_i includes an atom A_j , a bidirectional edge $v_{S_i} \leftrightarrow v_{A_j}$ will be established between the substructure-level node v_{S_i} and the atomic-level node v_{A_j} . This process is repeated for all relevant substructure-level and atomic-level nodes, as shown in the middle

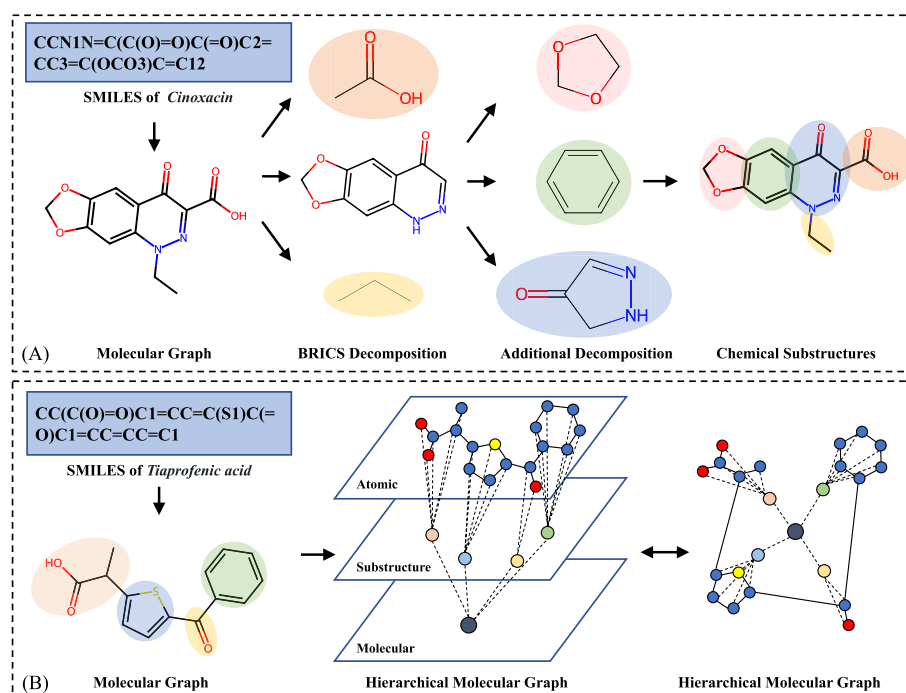


Fig. 2 The flowchart of the substructure extraction module in HDN-DDI. **A** The procedure of drug molecule decomposition. **B** The process for constructing a hierarchical molecular graph

part of Fig. 2B. This approach explicitly records the relationships between atoms and substructures, allowing for the integration of structural information for subsequent molecular encoding.

Furthermore, we introduce a molecular-level node v_M and establish bidirectional edges between this node and all substructure-level nodes. Specifically, for each substructure-level node v_{S_i} , a bidirectional edge $v_M \leftrightarrow v_{S_i}$ will be added, facilitating both the aggregation of local information and the dissemination of global information.

Ultimately, we construct a hierarchical molecular graph that includes all atomic-level nodes, which represent individual atoms; all substructure-level nodes, which correspond to the decomposed chemical substructures; and a molecular-level node, which represents the entire molecule. Additionally, the hierarchical molecular graph retains the original edges between atomic-level nodes, representing chemical bonds, along with all added bidirectional edges, which reflect relationships across different levels. In accordance with the original method [29], there are no edges between substructure-level nodes within a molecule.

In hierarchical molecular graphs, nodes at different levels can perceive information from other levels. Specifically, atomic-level nodes gather local information while molecular-level nodes aggregate global information. Furthermore, the substructure-level nodes learn the information of functional groups or substructures. Additionally, the atomic-level nodes can acquire information about the entire molecule and the molecular-level nodes can integrate and summarize all atomic information. Throughout this process, the substructure-level nodes play a pivotal role in facilitating the transmission of information between atomic-level nodes and molecular-level nodes.

HDN encoder

As depicted in Fig. 1B, the HDN Encoder comprises a series of HDN Blocks and each HDN Block consists of a hierarchical-view layer for individual drugs, an interactive-view layer for drug pairs and an update layer for information aggregation. Within both the hierarchical-view layer and the update layer, nodes can aggregate information from all their neighboring nodes, regardless of their levels. For instance, an atomic-level node can gather substructure information from neighboring substructure-level nodes while concurrently aggregating data from adjacent atomic-level nodes. Ultimately, a single-layer GAT is utilized to facilitate message transmission. The representation $h_i^{(l+1)}$ of the node i at the $(l + 1)$ -th block is calculated as illustrated from Equation (1) to Equation (3).

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}_i \cup \{i\}} \alpha_{ij} W^{(l+1)} h_j^{(l)} + b^{(l+1)} \right) \quad (1)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in \mathcal{N}_i \cup \{i\}} \exp(\text{LeakyReLU}(e_{ik}))} \quad (2)$$

$$e_{ij} = a^{(l+1)\text{T}} \left[W^{(l+1)} h_i^{(l)} \parallel W^{(l+1)} h_j^{(l)} \right] \quad (3)$$

where $W^{(l+1)} \in \mathbb{R}^{d^{(l+1)} \times d^{(l)}}$ and $a^{(l+1)} \in \mathbb{R}^{d^{(l+1)}}$ represent the trainable weights of the GAT in $(l + 1)$ -th block and $b^{(l+1)} \in \mathbb{R}^{d^{(l+1)}}$ denotes the trainable bias. T signifies

transpose operation. LeakyReLU($\cdot$) [34] and sigmoid function $\sigma(\cdot)$ are the activation functions. $d^{(l)}$ and $d^{(l+1)}$ are the input and output feature dimensions of the layer in $(l + 1)$ -th block respectively. $\mathcal{N}_i$ represents the set of all neighbors of the node i . $h_j^{(l)}$ denotes the representation of the node j at the l -th block and α_{ij} denotes the attention coefficient between the nodes i and j .

In the interactive-view layer, the bipartite graph $\tilde{G} = V_s^{(x)} \times V_s^{(y)}$ serves as the input where $V_s^{(x)}$ and $V_s^{(y)}$ represent the sets of substructure-level nodes in the hierarchical molecular graphs G_x and G_y respectively. Like the hierarchical-view layer, the interactive-view layer captures information from interactions between substructures using a single-layer GAT. Notably, the rationale for applying GAT in both layers is well-supported by previous studies [26, 27]. Although the GATs in these layers follow similar computational steps, they employ different weights, allowing each layer to capture distinct structural and interaction-specific characteristics.

Given that DDIs typically arise from a limited number of key substructures [23, 25], connections within the bipartite graph are randomly dropped during the encoding process. The computation steps for acquiring interaction representation $\tilde{h}_i^{(l+1)}$ of the node i at the $(l + 1)$ -th block are detailed from Equation (4) to Equation (6).

$$\tilde{h}_i^{(l+1)} = \sigma \left(\sum_{j \in \tilde{\mathcal{N}}_i \cup \{i\}} \beta_{ij} \tilde{W}^{(l+1)} h_j^{(l)} + \tilde{b}^{(l+1)} \right) \quad (4)$$

$$\beta_{ij} = \frac{\exp(\text{LeakyReLU}(\tilde{e}_{ij}))}{\sum_{k \in \tilde{\mathcal{N}}_i \cup \{i\}} \exp(\text{LeakyReLU}(\tilde{e}_{ik}))} \quad (5)$$

$$\tilde{e}_{ij} = \tilde{a}^{(l+1)\text{T}} \left[\tilde{W}^{(l+1)} h_i^{(l)} \parallel \tilde{W}^{(l+1)} h_j^{(l)} \right] \quad (6)$$

where $\tilde{W}^{(l+1)} \in \mathbb{R}^{d^{(l+1)} \times d^{(l)}}$ and $\tilde{a}^{(l+1)} \in \mathbb{R}^{d^{(l+1)}}$ represent the trainable weights of the GAT in the $(l + 1)$ -th block while $\tilde{b}^{(l+1)} \in \mathbb{R}^{d^{(l+1)}}$ denotes the trainable bias. $\tilde{\mathcal{N}}_i$ refers to the set of neighboring nodes of the node i in the bipartite graph after random connection pruning. Specifically, if the substructure-level node i is from the drug D_x , $\tilde{\mathcal{N}}_i$ represents a subset of substructure-level nodes from the drug D_y and vice versa. β_{ij} denotes the attention coefficient between the nodes i and j .

Unlike existing methods that consider all atomic nodes in the molecules [26, 27], HDN-DDI constructs bipartite graphs exclusively comprising substructure-level nodes. This novel approach prioritizes interaction information between substructures while filtering out noise from atomic-level nodes. Moreover, strategically dropping connections enables substructure-level nodes to capture diverse scopes of interest, which effectively prevents the homogenization of node representations during encoding. Furthermore, it broadens the scope of interest of identical nodes across different depths, which significantly enhances the effectiveness of deep networks. Consequently, this refined methodology yields an improved dual-view learning framework that aligns with the underlying mechanisms of DDIs and enhances the model's capability to identify key chemical substructures.

To maintain consistency in the representation dimensions of nodes, a nonlinear transformation is applied to all atomic-level and molecular-level nodes as illustrated in Eq. (7). It's worth noting that this transformation shares the same weight with the GAT in the interactive-view layer.

$$\tilde{h}_k^{(l+1)} = \sigma \left(\widetilde{W}^{(l+1)} h_k^{(l)} + \tilde{b}^{(l+1)} \right) \quad (7)$$

where $\tilde{h}_k^{(l+1)}$ is the interaction representation of the node k .

Given the hierarchical representation $h_i^{(l+1)}$ and interaction representation $\tilde{h}_i^{(l+1)}$ of the node i , the update layer aggregates them through a nonlinear layer as shown in Equation (8). Subsequently, it updates the entire graph through a single-layer GAT. The updated representations of all nodes serve as the inputs for the next HDN Block.

$$h_i^{(l+1)} = \text{ELU} \left(\text{MLP} \left(\left[h_i^{(l+1)} \parallel \tilde{h}_i^{(l+1)} \right] \right) \right) \quad (8)$$

where $\text{ELU}(\cdot)$ is a nonlinear activation function.

As described in Sect. 2.4, the molecular-level nodes are capable to aggregate information from all substructure-level nodes during encoding. Consequently, the hierarchical molecular graphs can directly obtain global molecular representations through these molecular-level nodes as illustrated in Equation (9).

$$g_x^{(l+1)} = h_{x_mol}^{(l+1)}, g_y^{(l+1)} = h_{y_mol}^{(l+1)} \quad (9)$$

where $h_{x_mol}^{(l+1)}$ and $h_{y_mol}^{(l+1)}$ are the respective representations of molecular-level nodes within the hierarchical molecular graphs G_x and G_y . Additionally, $g_x^{(l+1)}$ and $g_y^{(l+1)}$ denote the global representations of the drugs D_x and D_y in the $(l + 1)$ -th block. Compared to traditional graphs, hierarchical molecular graphs eliminate the need for additional pooling operations which effectively reduces computational costs.

HDN decoder

By iteratively learning with the HDN Blocks, the HDN Encoder effectively acquires multi-depth molecular representations that incorporate both hierarchical structure and interaction information of drugs. Subsequently, the HDN Decoder employs a co-attention mechanism [23, 26] to condense these representations. Specifically, the HDN Decoder initially assesses the layer-wise association strength between the global representations of two drugs across various depths as depicted in Eq. (10). It then learns the embedding representations of different interaction types and calculates the interaction scores between the two drugs at different depths. Finally, it combines the association matrix and the global representations at varying depths through point-wise multiplications to compute the DDI probability, as illustrated in Eq. (11).

$$\gamma_{ll'} = \alpha^T \tanh \left(W_x g_x^{(l)} + W_y g_y^{(l')} \right) \quad (10)$$

$$s(D_x, r, D_y) = \sigma \left(\sum_{l=1}^L \sum_{l'=1}^L \gamma_{ll'} g_x^{(l)} M_r g_y^{(l')} \right) \quad (11)$$

where $M_r \in \mathbb{R}^{d \times d}$ represents the embedding matrix of the interaction r . Moreover, $\alpha \in \mathbb{R}^d$ is a learnable weight vector and $W_x, W_y \in \mathbb{R}^{d \times d}$ denote learnable weight matrices. L is the number of the HDN Blocks while d is the dimension of molecular representation. $g_x^{(l)}, g_y^{(l)} \in \mathbb{R}^d$ denote the global representations of the drugs D_x and D_y in the l -th block respectively. σ is the sigmoid function and $s(D_x, r, D_y)$ stands for the probability of the interaction r occurring between the drugs D_x and D_y .

Loss function

As mentioned in Sect. 2.1, DDI prediction aims to predict whether the tuple (D_x, r, D_y) exists or not so it can be regarded as a binary classification problem. Since there are only positive samples in the dataset, we generate negative samples by replacing drugs [23, 26]. Specifically, a negative sample (D_x, r, D'_y) or (D'_x, r, D_y) is generated from a positive sample $t = (D_x, r, D_y)$ by replacing D_x with D'_x or D_y with D'_y . Note that there are no existing records of interactions between D'_x and D_y as well as D_x and D'_y before generation. After generation, there is an equal number of positive and negative samples in the dataset. Finally, the model is trained end-to-end with a binary cross-entropy as shown in Equation (12).

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|} \sum_{t=(D_x, r, D_y) \in \mathcal{T}} (\log p_t + \log (1 - p_t)) \quad (12)$$

where p_t is the existence probability of the tuple t which equals to $s(D_x, r, D_y)$ in Equation (11) and $\mathcal{T}$ denotes the set of both positive and negative samples.

Results

Datasets

We conducted experiments on two widely-used datasets: DrugBank [35] and Twosides [36]. DrugBank comprises 1,706 drugs with 86 types of interactions, yielding 191,808 positive DDI instances. Twosides is derived by Zitnik et al. [36] from the original dataset [37] and it includes 645 drugs with 963 types of interactions and 4,576,287 positive DDI instances. Both datasets lack negative DDI instances and to our knowledge there is currently no other database containing negative ones [17, 18], hence we generated negative samples in the same manner as in previous researches [25–27].

In addition to the size of the dataset, the two datasets differ in the meaning and number of drug-drug interaction types. DrugBank contains a single interaction between two drugs that describes how one drug affects the metabolism of another. For example, #Drug1 may increase the sedative activities of #Drug2 [35]. In contrast, Twosides has multiple interactions between two drugs that describe potential side effects. For example, the co-administration of #Drug1 and #Drug2 may result in headaches and fever [25]. Therefore, the two datasets could not be merged and were used separately for evaluation.

Finally, we conducted evaluations in two settings: warm-start and cold-start. In the warm-start setting, a drug might appear in both the training and testing sets and the

split ratio for training, validation and testing in DrugBank and Twosides was 6:2:2. In the cold-start setting, each drug would be exclusively assigned to either the training or testing set so that there was no overlap between drugs in the two sets. Following previous works [23–27], we used only the DrugBank dataset in the cold-start setting due to its larger number of drugs, totaling 1706. Additionally, we randomly selected 20% of the drugs as unseen for testing and another 5% for validation. Both the warm-start and cold-start settings employ split ratios that align with those found in existing studies [25–27].

The scale of positive samples in the datasets is detailed in the Table 1. In the cold-start setting, the evaluation was conducted in three times and the testing sets were further split into S1 and S2 partitions, which will be introduced in Sect. 3.6. It’s worth notable that the final sample size used for evaluation is twice the number of positive samples, as the number of generated negative samples is equal to the number of original positive samples.

Parameters

As shown in Fig. 1, HDN-DDI consists of the HDN Encoder and the HDN Decoder with the former comprising multiple HDN Blocks. Each HDN Block contains a hierarchical-view layer, an interactive-view layer and an update layer and all of layers employ a multi-head GAT for message transmission. The optimal hyperparameters are determined through random search on the validation set and the search space is detailed in Additional file 2. Specifically, HDN-DDI contains six HDN Blocks. In each block, both the hierarchical-view and interactive-view layers produce a 64-dimensional representation while the update layer generates a 128-dimensional representation. All GATs in HDN-DDI utilize two attention heads. Additionally, the batch size used during model training is 512 and the learning rate is set to 0.001.

Evaluation metrics

Due to a single evaluation metric is often insufficient to effectively and comprehensively reflect the performance of the models [17, 38, 39], we evaluated HDN-DDI and baselines with four metrics which are widely-used in previous studies [23–27]. These metrics are listed as follows:

- Accuracy (ACC): it’s the proportion of correct predictions in all predictions.

Table 1 The scale of positive samples in the warm-start and cold-start setting

Setting	Warm-start setting		Cold-start setting		
Dataset	DrugBank	Twosides	DrugBank Fold 0	DrugBank Fold 1	DrugBank Fold 2
Training set	115,085	2,745,772	108,396	114,617	110,094
Validation set	38,361	915,257	15,388	11,733	15,217
Testing set	38,362	915,258	68,024	65,458	66,497
- S1 partition	–	–	7,369	6,719	6,935
- S2 partition	–	–	60,655	58,739	59,562
Total	191,808	4,576,287	191,808	191,808	191,808

The symbol “–” indicates that there were no S1 and S2 partitions in the warm-start setting

- Area under the receiver operating characteristic curve (AUROC): it equals to the probability that the model scores a randomly selected positive example higher than a randomly selected negative one [26].
- Area under the precision-recall curve (AUPR): it reflects the model's ability to maintain high precision and high recall. In the case where positive samples are more important, AUPR is a more suitable metric [39, 40].
- F1 score (F1): it's the harmonic mean of precision and recall. This metrics also intuitively reflects the model's prediction reliability for positive samples [41].

In this work, the original datasets contain only positive samples while negative samples are generated through replacing drugs hence we focused on the model's prediction of positive samples. Furthermore, the DDI predictions is regarded a binary classifying problem and both AUPR and F1 score can reflect the model's performance on positive samples, so we focused more on these two metrics.

Baselines

We compared HDN-DDI with the current state-of-the-art in both warm-start and cold-start settings and the baselines include both substructure-based algorithms and dual-view representation learning algorithms:

- MHCADDI [20]: utilizes a co-attention mechanism to integrate the joint information of drug pairs into the representation learning for individual drugs.
- SSI-DDI [23]: employs a multi-layer GAT to extract substructures and calculates the probability of interactions between substructures to predict DDIs.
- SA-DDI [24]: employs a substructure-aware GNN to capture size-adaptive substructures and predicts DDI through a novel substructure attention mechanism.
- GMPNN-CS [25]: learns the substructure information across various scales and models the interactions between substructures for DDI prediction.
- DSN-DDI [26]: employs local and global representation learning modules iteratively and learns substructures from respective drugs and drug pair simultaneously.
- BDN-DDI [27]: captures pairwise atomic interactions through an interactive GNN and learns substructures through dual-view representation learning.

Results in warm-start setting

In the warm-start setting, there is an overlap between drugs in the training and testing sets. Additionally, we conducted three runs and reported the average as the final performance. In each run, the dataset was randomly stratified to split training, validation and testing sets with an equal ratio of interaction types. For a fair comparison, negative samples were generated and datasets were split before training so that all models were trained and tested on the same data.

The average performance of all models over the three runs is shown in Table 2. It can be observed that HDN-DDI outperformed all baselines on both DrugBank and Twosides across all evaluation metrics. While the current state-of-the-art has achieved impressive accuracies of 96.94% and 99.07% on these two datasets, HDN-DDI surpassed them

Table 2 The performance of HDN-DDI and baselines on two datasets in the warm-start setting (%)

Model	DrugBank				Twosides			
	ACC	AUROC	AUPR	F1	ACC	AUROC	AUPR	F1
MHCADDI	83.80	91.16	89.26	85.06	–	88.20	–	–
SSI-DDI	96.33	98.95	98.57	96.38	78.20	85.85	82.71	79.81
GMPNN-CS	95.30	98.46	97.94	95.39	82.83	90.07	87.24	84.08
SA-DDI	96.23	98.80	98.36	96.29	87.45	93.17	90.57	88.35
DSN-DDI	96.94	99.47	99.37	96.93	98.83	99.90	99.89	98.83
BDN-DDI	96.33	99.34	99.28	94.40	99.07	99.92	99.90	99.07
HDN-DDI	97.93	99.73	99.72	98.03	99.38	99.97	99.97	99.38
Improvement	0.99	0.26	0.35	1.10	0.31	0.05	0.07	0.31

The highest value in each column is shown in bold

with even higher accuracies of 97.93% on DrugBank and 99.38% on Twosides. Moreover, HDN-DDI achieved a superior AUPR of 99.72% on DrugBank and 99.97% on Twosides, underscoring its outstanding predictive ability for positive samples. These results demonstrate that HDN-DDI excels in predicting DDIs involving existing drugs with remarkable metrics.

Results in cold-start setting

In the cold-start setting, there is no overlap between the drugs in the training and testing sets. In other words, the training and testing sets are partitioned by drugs. The cold-start setting can be used to evaluate the models' ability to identify DDIs involving previously unseen drugs. Since the models lack prior structural information about the drugs in the testing set, predicting DDIs in this setting becomes more challenging and it requires models to have better generalization capabilities [26, 42]. Formally, let $\mathcal{G}$ denote the set of all drugs, $\mathcal{G}_{new}$ represent the set of new drugs and $\mathcal{G}_{old}$ denote the set of drugs used for training. Evidently, $\mathcal{G}_{new} \cup \mathcal{G}_{old} = \mathcal{G}$ and $\mathcal{G}_{new} \cap \mathcal{G}_{old} = \emptyset$. In the cold-start setting, the dataset is partitioned into the following parts:

$$\begin{aligned}
 \mathcal{D}_{train_val} &= \{(D_x, r, D_y) | D_x \in \mathcal{G}_{old} \wedge D_y \in \mathcal{G}_{old}\} \\
 \mathcal{D}_{s1} &= \{(D_x, r, D_y) | D_x \in \mathcal{G}_{new} \wedge D_y \in \mathcal{G}_{new}\} \\
 \mathcal{D}_{s2} &= \{(D_x, r, D_y) | D_x \in \mathcal{G}_{new} \wedge D_y \in \mathcal{G}_{old}\} \\
 &\quad \cup \{(D_x, r, D_y) | D_x \in \mathcal{G}_{old} \wedge D_y \in \mathcal{G}_{new}\}
 \end{aligned}$$

where $\mathcal{D}_{train_val}$ is the set which contains tuples only involving seen drugs while $\mathcal{D}_{s1}$ and $\mathcal{D}_{s2}$ are testing sets that include tuples consisting of at least one new drug. Moreover, the training and validation set will be split from $\mathcal{D}_{train_val}$ in the same way.

Finally, we repeated three times and reported the average performance. In each run, we randomly sampled 20% drugs as new drugs to construct different testing sets across three runs. Notably, negative sample generation on the training and testing sets was conducted based on their own contained drugs so that the generated samples align with the condition of the cold-start setting. Both drug selection and negative sample generation were conducted before training to ensure that all models shared the same training, validation and testing sets.

The average performance of all models over three runs is presented in Table 3 and it can be observed that the cold-start setting significantly impacted the performance of all models. However, HDN-DDI consistently outperformed the baseline models with remarkable enhancements. Compared with the current state-of-the-art, the accuracy of HDN-DDI improved by 4.96% on $\mathcal{D}_{s1}$ and 4.08% on $\mathcal{D}_{s2}$ and its F1 score enhanced by 7.08% on $\mathcal{D}_{s1}$ and 5% on $\mathcal{D}_{s2}$. Although the F1 score of the BDN-DDI's wo_bilinear variant reached 71.30% [27], HDN-DDI still led with a 6.12% advantage. Moreover, the AUROC and AUPR of HDN-DDI also demonstrated improvement varying 2.24% to 3.81%. These results suggested that HDN-DDI can predict DDIs involving new drugs with greater accuracy. Although the negative samples in the datasets are artificially generated, HDN-DDI was still able to learn well and showed greater predictive ability for original positive samples. In conclusion, HDN-DDI has achieved the state-of-the-art performance in both the warm-start and cold-start settings.

Ablation study

To further explore the contribution of each module in HDN-DDI, we conducted an ablation study in the cold-start setting which better distinguished the performance of models. The variants utilized in this study are as follows:

- wo-SEM: a model that omits the substructure extraction module and learns only from traditional molecular graphs, thus validating the effectiveness of the module in molecular representation learning and DDI prediction.
- wo-EDV: a model that replaces the enhanced dual-view representation learning method with the initial method [26], to confirm the superior capability of the former in capturing interaction information.
- with-pool: a model that acquires global representations through a simple sum function without distinguishing the contribution of substructures, to underscore the significance of capturing key substructures in molecules.
- with-SAG: a model that generates global representations through the widely-used method SAGPooling [43] rather than from the molecular-level nodes, thus verifying the latter's effectiveness in global representation learning.

Table 3 The performance of HDN-DDI and baselines on DrugBank in the cold-start setting (%)

Model	$\mathcal{D}_{s1}$ (new drug ↔ new drug)				$\mathcal{D}_{s2}$ (new drug ↔ old drug)			
	ACC	AUROC	AUPR	F1	ACC	AUROC	AUPR	F1
MHCADDI	66.50	72.53	71.06	67.21	70.58	77.84	76.16	72.74
SSI-DDI	65.40	73.43	75.03	54.12	76.38	84.23	84.94	73.54
GMPNN-CS	68.57	74.96	75.44	65.32	77.72	84.84	84.87	78.29
SA-DDI	67.15	73.62	73.39	63.40	75.55	82.95	84.11	71.94
DSN-DDI	73.42	81.79	81.82	70.34	81.92	91.01	91.09	80.18
BDN-DDI	74.88	86.10	85.84	70.21	85.35	93.68	93.64	84.34
HDN-DDI	79.84	89.48	89.65	77.42	89.43	95.92	95.95	89.34
Improvement	4.96	3.38	3.81	7.08	4.08	2.24	2.31	5.00

The highest value in each column is shown in bold

As shown in Table 4, HDN-DDI has surpassed all its variants, which highlights the importance of each module for improving model performance. A detailed analysis of the ablation experiment results reveals the following insights:

- (1) The substructure extraction module plays a crucial role in HDN-DDI, because there was a significant performance drop in wo-SEM while the variants retaining this module outperformed all baselines. This underscores the effectiveness of the module in enhancing drug molecular representation learning by integrating information from real chemical substructure and hierarchical structure in molecules.
- (2) The enhanced dual-view learning method proposed in this work exhibited superior efficacy in encoding drug pairs than the initial method [26], which yielded accuracy enhancements of 3.69% and 4.45% on two partitions. This highlights the effectiveness of the enhanced method in improving dual-view representation learning.
- (3) The method of extracting global representation from molecular-level nodes surpassed pooling methods such as sum pooling and SAGPooling while reduced computational costs. This finding further substantiates the superiority of hierarchical molecular graphs in molecular representation learning.

In summary, the ablation study underscores that modules within HDN-DDI such as hierarchical molecular graphs and the enhanced dual-view representation learning method significantly improve the model’s learning capabilities.

Real-world applications

To evaluate the effectiveness of HDN-DDI in real-world scenarios, we trained the model using existing drugs and predicted DDIs for new FDA-approved drugs [44]. For each drug molecule in DrugBank, we collected all FDA-approved drugs containing it as the active ingredient and identified the earliest approval date among these drugs as its own approval date. Subsequently, we categorized the drugs into two groups based on their approval dates: existing drugs (approved before 2015) and new drugs (approved in 2015 or later). The former group comprised 1,585 drugs, while the latter consisted of 121 drugs. Ultimately, the training dataset comprised 166,342 positive DDI tuples involving only existing drugs, while the test dataset comprised 25,466 positive DDI tuples involving at least one newly approved drug.

Since DSN-DDI [26] and BDN-DDI [27] have demonstrated superior performance among all baseline models, our study focused on comparing HDN-DDI with these two

Table 4 The performance of HDN-DDI and its variants on DrugBank in the cold-start setting (%)

Model	$\mathcal{D}_{s1}$ (new drug ↔ new drug)				$\mathcal{D}_{s2}$ (new drug ↔ old drug)			
	ACC	AUROC	AUPR	F1	ACC	AUROC	AUPR	F1
wo-SEM	69.96	77.23	77.62	66.83	78.01	85.80	85.79	78.15
wo-EDV	76.15	85.44	84.91	74.00	84.98	92.60	92.17	84.97
with-pool	78.56	87.56	87.33	76.89	87.18	94.42	94.22	87.24
with-SAG	78.20	88.05	88.15	75.41	87.67	94.71	94.70	87.60
HDN-DDI	79.84	89.48	89.65	77.42	89.43	95.92	95.95	89.34

The highest value in each column is shown in bold.

models. To ensure fairness, we evaluated DSN-DDI and its variant with six encoding blocks, which aligned with the structure of HDN-DDI and BDN-DDI that comprised six encoding blocks. Hyperparameters for all models were initialized according to the default values specified in their respective original papers. The results presented in Additional file 3 underscore HDN-DDI's significant superiority across all evaluation metrics, with notable enhancements of 5.86% in accuracy, 4.62% in AUPR and 6.09% in F1 score. Moreover, as shown in Fig. 3, the area under the curves (i.e. AUROC and AUPR) of HDN-DDI substantially surpassed those of the other models. This observation remains consistent in the cold-start setting and it highlights HDN-DDI's robust generalization capability and promising applicability in real-world scenarios.

Case study

To verify the rationality of HDN-DDI for distinguishing key substructures, we conducted visualizations of the prediction cases as shown in Fig. 4. Within each HDN Block, the GAT in the update layer calculates and outputs the contribution of each substructure in the molecules. For HDN-DDI with multiple HDN Blocks, we averaged the contribution output of each block to obtain an average contribution value for each substructure. Subsequently, we selected the substructures with the top 40% contribution value as the extracted key substructures and analyzed these cases from the perspective of drugs' characteristics and interaction mechanisms:

(1) Amyl Nitrite and Vardenafil. The former is an organic nitrate commonly used to alleviate angina pectoris and the latter is a Phosphodiesterase-5 inhibitor employed for erectile dysfunction treatment. Both drugs have antihypertensive effects and their concurrent usage can precipitate a swift reduction in blood pressure [35, 45]. Specifically, Vardenafil serves as a substrate of P-glycoprotein (P-gp) while Amyl Nitrite is an inhibitor of P-gp activity. Their co-administration enhances the bioavailability of Vardenafil which potentially amplifies its therapeutic efficacy [27, 45]. The efficacy of Amyl Nitrite predominantly hinges on its nitroso functional group while Vardenafil's primary functional groups include the pyridine ring and amide bond [27]. The same is true for Erythrityl tetranitrate and Tadalafil [35]. As shown in (a) and (b) as well as (c) and (d) of Fig. 4, HDN-DDI accurately identified these crucial functional groups.

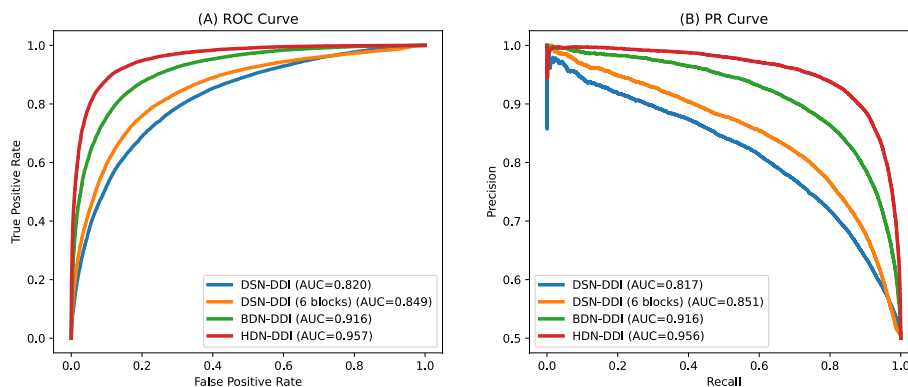


Fig. 3 The prediction curves of the HDN-DDI and baseline models for new FDA-approved drugs. **A** The receiver operating characteristic (ROC) curve of models on the testing set. **B** The precision-recall (PR) curve of models on the testing set. The AUC denotes area under the curve

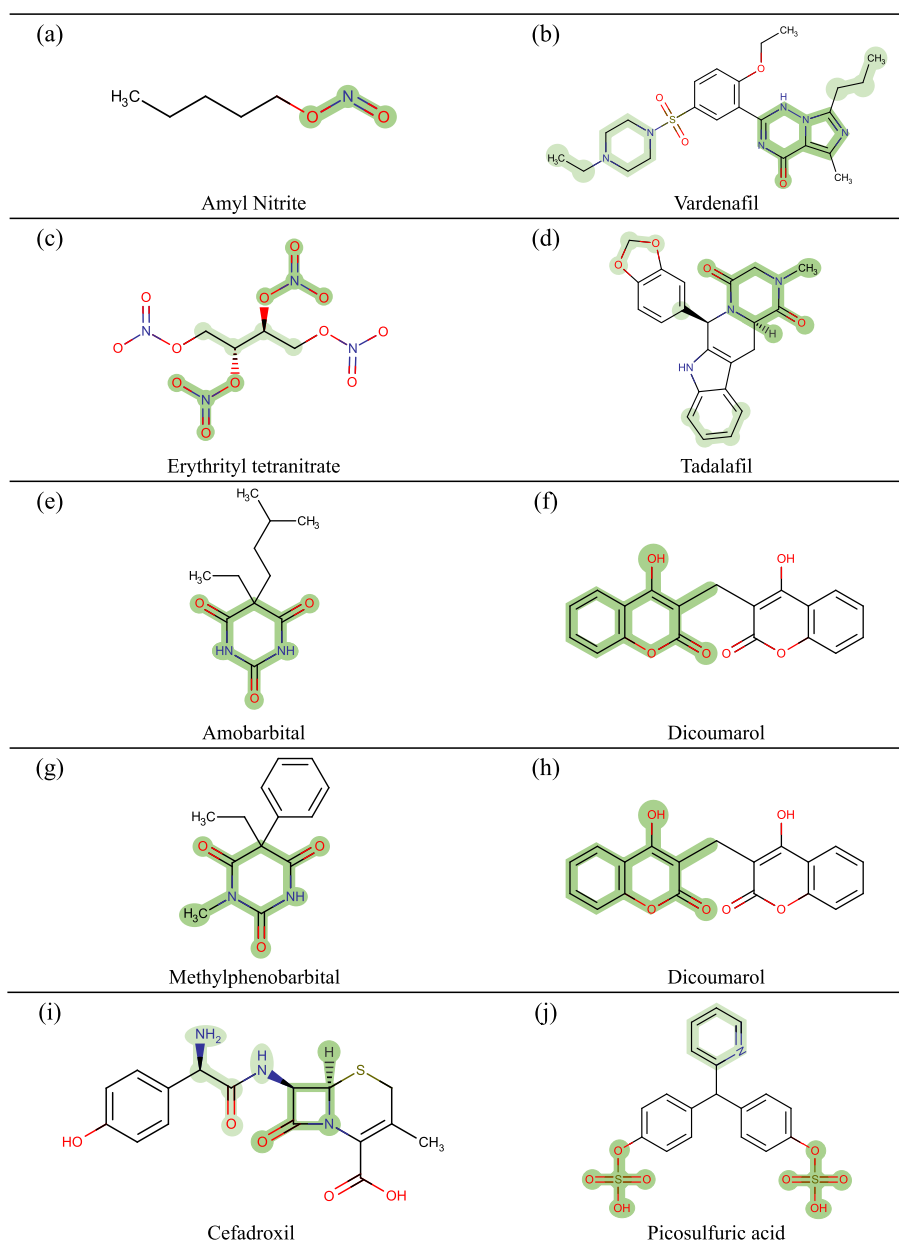


Fig. 4 The visualization of HDN-DDI for five groups of DDI prediction cases. The substructures that contribute to the top 40% are highlighted. A darker shade indicates a greater contribution of the corresponding substructure

(2) Amobarbital and Dicoumarol. The former is a barbiturate derivative commonly used for hypnosis and sedation while the latter is an anticoagulant agent. Their co-administration diminishes coagulation activity of Dicoumarol [35, 46]. Specifically, Amobarbital can enhance the activity of liver microsomal enzymes which expedites the metabolism of Dicoumarol and weakens its efficacy. A similar interaction is observed between Methylphenobarbital and Dicoumarol [46]. Barbituric acid emerges as the pivotal functional group among barbiturate derivatives while in Dicoumarol both the benzene ring and the furan ring play crucial roles [26, 47]. As shown in (e) and (f) as well

as (g) and (h) in Fig. 4, HDN-DDI accurately identifies these critical functional groups within the respective drugs.

(3) Cefadroxil and Picosulfuric acid. The former is a cephalosporin antibiotic utilized to treat bacterial infections with the cephalosporin β -lactam ring as the principal functional group [48]. The latter is a colon-cleansing laxative characterized primarily by its sulfonic acid group [23]. The co-administration of these drugs results in a reduction in the efficacy of Picosulfuric acid [23, 35]. As depicted in (i) and (j) of Fig. 4, HDN-DDI accurately identified the β -lactam ring and the sulfonic acid group. These findings underscore the capability of HDN-DDI to precisely pinpoint crucial substructures within drug molecules.

Discussion

Here, we propose an accurate and robust DDI prediction model, named HDN-DDI. Comprehensive experiments including warm-start and cold-start settings demonstrate that HDN-DDI exhibits superior prediction performance. Additionally, the ablation study further validates the importance of each module in HDN-DDI. By employing an explainable substructure extraction module to decompose drug molecules, HDN-DDI effectively harnesses information from real chemical substructures. By representing drug molecules with “atomic-substructure-molecular” hierarchical molecular graphs, HDN-DDI incorporates hierarchical structures that significantly enhance subsequent molecular representation learning. Furthermore, the enhanced dual-view representation learning focuses on interactions between chemical substructures rather than individual atoms, which reduces learning noise and helps the model identify chemical substructures that critically influence DDIs. As a result, HDN-DDI excels in predicting DDIs and demonstrates robust generalization ability for novel drugs.

While the HDN-DDI model shows promise, there are areas for improvement. First, some drug molecules may not undergo sufficient decomposition due to the constraints imposed by the current decomposition rules. Additionally, we observe that incorporating edge attributes leads to a decline in model performance, a trend also observed in DSN-DDI. We hypothesize that this is due to the initial GAT architecture not considering edge attributes during encoding. Effectively incorporating edge information requires a more complex network structure, which could increase the risk of over-fitting and reduce model performance. Moreover, the case study suggests that DDIs are also influenced by drug metabolism, underscoring the importance of combining external knowledge with chemical substructures for prediction.

Future research will focus on refining the decomposition rules and expanding the substructure library to enhance the model's ability to extract substructures from a broader range of drugs. Furthermore, incorporating external knowledge, utilizing edge attributes, and constructing more effective hierarchical molecular graphs will provide promising avenues for improvement.

Conclusion

In this study, we introduce a novel DDI prediction model that addresses challenges encountered in previous studies, such as the interpretability of identified substructures and the isolation of chemical substructures. HDN-DDI utilizes hierarchical

molecular graphs combined with an enhanced dual-view representation learning method, attaining state-of-the-art performance. Our experimental results on two widely-used datasets show that HDN-DDI achieves high accuracies of 97.93% and 99.38% in the warm-start setting. Furthermore, HDN-DDI exhibits notable enhancements with a 5% increase in accuracy and a 7% improvement in F1 score in the cold-start setting and real-world applications. These results highlight its superior generalization capability for previously unseen drugs, which is crucial for practical applications. With its integration of real chemical substructure information and its generalization ability for new drugs, HDN-DDI represents a significant advancement in DDI prediction.

In summary, by offering accurate predictions and robust generalization across various scenarios, HDN-DDI provides a valuable tool for identifying potential DDIs, promoting safer and more effective drug development and co-administration.

Abbreviations

DDI	Drug–drug interaction
GNN	Graph neural network
GAT	Graph attention network
ACC	Accuracy
AUROC	Area under the receiver operating characteristic
AUPR	Area under the precision–recall curve
F1	F1 score
ROC	Receiver operating characteristic
PR	Precision–recall
P-gp	P-glycoprotein

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-025-06052-0>.

Additional file

Supplementary file 1.
Supplementary file 2.
Supplementary file 3.

Acknowledgements

Not applicable.

Author Contributions

JS and HZ designed the study and drafted the manuscript. JS carried out models design, conducted the experiments and analysed the results. HZ arranged the study plan and reviewed the manuscript. Both authors read and approved the final manuscript.

Funding

This work was supported by the National Key Technologies R&D Program of China [2017YFA0505502] and the Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38000000]. The funders had no role in the design of the study; collection, analysis and interpretation of data; decision to publish; or preparation of the manuscript.

Availability of data and materials

The source code and figures for this paper are available online at <https://github.com/jcsun-00/HDN-DDI>. The DrugBank dataset can be accessed at <https://github.com/jcsun-00/DrugBank>. The Twosides dataset can be accessed at <https://github.com/jcsun-00/Twosides>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 26 May 2024 Accepted: 15 January 2025

Published online: 25 January 2025

References

- Lajoie AC, Lauzière G, Lega JC, Lacasse Y, Martin S, Simard S, et al. Combination therapy versus monotherapy for pulmonary arterial hypertension: a meta-analysis. *Lancet Respir Med*. 2016;4(4):291–305.
- Mokhtari RB, Homayouni TS, Baluch N, Morgatskaya E, Kumar S, Das B, et al. Combination therapy in combating cancer. *Oncotarget*. 2017;8(23):38022.
- Khawcharoenporn T, Chuncharunee A, Maluangnon C, Taweesakulvashra T, Tiamsak P. Active monotherapy and combination therapy for extensively drug-resistant *Pseudomonas aeruginosa* pneumonia. *Int J Antimicrob Agents*. 2018;52(6):828–34.
- Jin B, Yang H, Xiao C, Zhang P, Wei X, Wang F. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 31; 2017. p. 1367–1373.
- Poleksic A, Xie L. Database of adverse events associated with drugs and drug combinations. *Sci Rep*. 2019;9(1):20025.
- Malki MA, Pearson ER. Drug-drug-gene interactions and adverse drug reactions. *Pharmacogenomics J*. 2020;20(3):355–66.
- Pang K, Wan YW, Choi WT, Donehower LA, Sun J, Pant D, et al. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics*. 2014;30(10):1456–63.
- Safdari R, Ferdousi R, Aziziheris K, Niakan-Kalhor SR, Omid Y. Computerized techniques pave the way for drug-drug interaction prediction and interpretation. *Bioimpacts*. 2016;6(2):71–8.
- Madani Tonekaboni SA, Soltan Ghoraiie L, Manem VSK, Haibe-Kains B. Predictive approaches for drug combination discovery in cancer. *Brief Bioinform*. 2018;19(2):263–76.
- Yu H, Mao KT, Shi JY, Huang H, Chen Z, Dong K, et al. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol*. 2018;12:101–10.
- Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PLoS ONE*. 2018;13(5): e0196865.
- Shi JY, Mao KT, Yu H, Yiu SM. Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization. *J Cheminform*. 2019;11:1–16.
- Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*. 2020;36(15):4316–22.
- Feng YH, Zhang SW. Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs. *Molecules*. 2022;27(9):3004.
- Yin Q, Fan R, Cao X, Liu Q, Jiang R, Zeng W. Deepdrug: a general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. *Quantit Biol*. 2023;11(3):260–74.
- Vo TH, Nguyen NTK, Kha QH, Le NQK. On the road to explainable AI in drug-drug interactions prediction: a systematic review. *Comput Struct Biotechnol J*. 2022;20:2112–23.
- Wu L, Wen Y, Leng D, Zhang Q, Dai C, Wang Z, et al. Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform*. 2022;23(1):bbab355.
- Lin X, Dai L, Zhou Y, Yu ZG, Zhang W, Shi JY, et al. Comprehensive evaluation of deep and graph learning on drug-drug interactions prediction. *Brief Bioinform*. 2023;24(4):bbad235.
- Xu N, Wang P, Chen L, Tao J, Zhao J. MR-GNN: multi-resolution and dual graph neural network for predicting structured entity interactions. In: *Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI-19*; 2019. p. 3968–3974.
- Deac A, Huang YH, Veličković P, Liò P, Tang J. Drug-drug adverse effect prediction with graph co-attention. *ArXiv* 2019 May; [arxiv:abs/1905.00534](https://arxiv.org/abs/1905.00534).
- Lin J, Wu L, Zhu J, Liang X, Xia Y, Xie S, et al. R2-DDI: relation-aware feature refinement for drug-drug interaction prediction. *Brief Bioinform*. 2023;24(1):bbac576.
- Harrold M, Zavod R. Basic concepts in medicinal chemistry. ASHP; 2013.
- Nyamabo AK, Yu H, Shi JY. SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction. *Brief Bioinform*. 2021;22(6):bbab133.
- Yang Z, Zhong W, Lv Q, Chen CYC. Learning size-adaptive molecular substructures for explainable drug-drug interaction prediction by substructure-aware graph neural network. *Chem Sci*. 2022;13(29):8693–703.
- Nyamabo AK, Yu H, Liu Z, Shi JY. Drug-drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinform*. 2022;23(1):bbab441.
- Li Z, Zhu S, Shao B, Zeng X, Wang T, Liu TY. DSN-DDI: an accurate and generalized framework for drug-drug interaction prediction by dual-view representation learning. *Brief Bioinform*. 2023;24(1):bbac597.
- Ning G, Sun Y, Ling J, Chen J, He J. BDN-DDI: a bilinear dual-view representation learning framework for drug-drug interaction prediction. *Comput Biol Med*. 2023;165: 107340.
- Zhang Z, Liu Q, Wang H, Lu C, Lee CK. Motif-based graph self-supervised learning for molecular property prediction. *Adv Neural Inf Process Syst*. 2021;34:15870–82.
- Zang X, Zhao X, Tang B. Hierarchical molecular graph self-supervised learning for property prediction. *Commun Chem*. 2023;6(1):34.
- 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.
- Fan Z, Zheng H. Pretraining molecular and substructural encoders for predicting drug-drug interactions in cold-start scenarios. In: *Chen M, Ning G, (Eds) Second International Conference on Biomedical and Intelligent Systems (IC-BIS 2023)*. vol. 12724. SPIE; 2023. p. 127240V–127240V–9.

32. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. 2017 Oct; [arxiv:1710.10903](#).
33. Landrum G, et al. RDKit: a software suite for cheminformatics, computational chemistry, and predictive modeling. Greg Landrum. 2013;8:5281.
34. Maas AL, Hannun AY, Ng AY et al. Rectifier nonlinearities improve neural network acoustic models. In: Proc. icml. vol. 30; 2013. p. 3.
35. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074–82.
36. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457–66.
37. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012;4(125):125ra31–125ra31.
38. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning; 2006. p. 233–240.
39. McDermott M, Hansen LH, Zhang H, Angelotti G, Gallifant J. A closer look at auroc and auprc under class imbalance. *ArXiv* 2024 Jan.
40. Zhou T. Discriminating abilities of threshold-free evaluation metrics in link prediction. *Physica A*. 2023;615: 128529.
41. Yang Y, Lichtenwalter RN, Chawla NV. Evaluating link prediction methods. *Knowl Inf Syst*. 2015;45:751–82.
42. Dewulf P, Stock M, De Baets B. Cold-start problems in data-driven prediction of drug-drug interaction effects. *Pharmaceuticals*. 2021;14(5):429.
43. Lee J, Lee I, Kang J. Self-attention graph pooling. In: International conference on machine learning; 2019. p. 3734–3743.
44. Mullard A. 2020 FDA drug approvals. *Nat Rev Drug Discov*. 2021;20(2):85–91.
45. Slavin S. Recreational use of amyl nitrite. *Venereology*. 2001;14(2).
46. Hudson RJ, Stanski DR. Barbiturates - pharmacokinetics and pharmacodynamics. *Clin Anaesthesiol*. 1984;2(1):27–41.
47. Coupey SM. Barbiturates. *Pediatr Rev*. 1997;18(8):260–5.
48. de Marco BA, Salgado HRN. Characteristics, properties and analytical methods of cefadroxil: a review. *Crit Rev Anal Chem*. 2017;47(2):93–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.