

## Data and text mining

# H<sup>2</sup>GnnDTI: hierarchical heterogeneous graph neural networks for drug–target interaction prediction

Yueying Jing<sup>1</sup>, Dongxue Zhang<sup>1</sup>, Limin Li<sup>1,\*</sup> 

<sup>1</sup>School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

\*Corresponding author. School of Mathematics and Statistics, Xi'an Jiaotong University, Xianning West 28, Xi'an, Shaanxi 710049, China.  
E-mail: liminli@mail.xjtu.edu.cn.

Associate Editor: Macha Nikolski

## Abstract

**Motivation:** Identifying drug–target interactions (DTIs) is a crucial step in drug repurposing and drug discovery. The significant increase in demand and the expensive nature for experimentally identifying DTIs necessitate computational tools for automated prediction and comprehension of DTIs. Despite recent advancements, current methods fail to fully leverage the hierarchical information in DTIs.

**Results:** Here, we introduce H<sup>2</sup>GnnDTI, a novel two-level hierarchical heterogeneous graph learning model to predict DTIs, by integrating the structures of drugs and proteins via a low-level view GNN and a high-level view GNN. The hierarchical graph consists of high-level heterogeneous nodes representing drugs and proteins, connected by edges representing known DTIs. Each drug or protein node is further detailed in a low-level graph, where nodes represent molecules within each drug or amino acids within each protein, accompanied by their respective chemical descriptors. Two distinct low-level graph neural networks are first deployed to capture structural and chemical features specific to drugs and proteins from these low-level graphs. Subsequently, a high-level graph encoder (GE) is used to comprehensively capture and merge interactive features pertaining to drugs and proteins from the high-level graph. The high-level encoder incorporates a structure and attribute information fusion module designed to explicitly integrate representations acquired from both a feature encoder and a GE, facilitating consensus representation learning. Extensive experiments conducted on three benchmark datasets have shown that our proposed H<sup>2</sup>GnnDTI model consistently outperforms state-of-the-art deep learning methods.

**Availability and implementation:** The codes are freely available at <https://github.com/LiminLi-xjtu/H2GnnDTI>.

## 1 Introduction

The identification of drug–target interactions (DTIs) is an important part of pharmacology and drug discovery. The traditional process of drug development is time-consuming and costly, whereas very few drugs ever make it to the clinic (Zhao *et al.* 2021). With the production of a large amount of biological activity data in recent years, predicting DTIs through deep learning technology has become a popular research topic. Many studies have adopted computational methods to effectively identify drug target associations on a large scale (Tsubaki *et al.* 2019, Wang *et al.* 2020). Recently, deep learning-based methods have shown good performance, but there are still two challenges: how to clearly model and learn local representations between drugs and targets for better prediction, and how to optimize the generalization performance of novel (unseen) proteins and drugs.

Over the past decades, many end-to-end models have been applied to study DTIs. DeepConv-DTI (Lee *et al.* 2019) utilizes multi-scale one-dimensional CNN (1D-CNN) layers to obtain protein features and get Extended Connectivity Fingerprints (ECFP) (Rogers and Hahn 2010) of drugs as input. To capture any relationship among atoms in a sequence, a Transformer-based DTI model (Kalakoti *et al.* 2022) uses multi-layered bidirectional Transformer encoders (Vaswani *et al.* 2017) to learn the high-dimensional structure of a

molecule from the simplified molecular-input line entry system (SMILES) string. In recent years, graph neural networks (GNNs) have achieved good performance in node/graph representation learning by simultaneously utilizing node attributes and graph high-levelology. For instance, Gu *et al.* (2022) proposed a heterogeneous GNN approach to integrate large-scale heterogeneous biological networks in the field of drug repositioning, which substantially improves prediction accuracy by deeply mining relationships between diverse biological entities. The DrugBAN (Bai *et al.* 2023) works on drug molecular graphs and target protein sequences to perform prediction, and proposes an interpretable bilinear attention network with domain adaptation to improve drug–target prediction accuracy.

While previous efforts have shown promising results by utilizing both types of information, these methods typically focus on a single view of either sequences or graphs. Few approaches effectively model the natural hierarchy of DTIs by integrating and connecting both views. Furthermore, existing methods often lack a suitable mechanism for information fusion of drug molecule graphs and protein residue graphs. Information from multiple sources is commonly aligned or concatenated, which limits the depth of interaction and integration of information.

To address the aforementioned challenges, we propose an end-to-end bio-inspired approach named H<sup>2</sup>GnnDTI for

predicting DTIs. The hierarchical graph model has previously been used in a homogeneous manner to investigate protein-protein interactions (Gao *et al.* 2023), demonstrating effective performance in capturing protein features for predicting homogeneous interactions. However, DTI prediction encompasses a more complicated task as it involves predicting heterogeneous interactions between drugs and targets, where the hierarchical knowledge of structural information plays a crucial role in understanding the molecular intricacies of DTIs. H<sup>2</sup>GnnDTI utilizes a hierarchical heterogeneous graph and incorporates both low-level and high-level GNNs (Kipf and Welling 2017, Xu *et al.* 2018) to capture structural and interactive features of drugs and proteins. This approach could address the limitations in previous prediction methods. On the one hand, high-level and low-level GNNs enable information propagation from a molecule in a drug graph or an amino acid in a protein graph to drugs or proteins in the DTI graph, and thus make full use of hierarchical structure to learn efficient representations of drug and protein structures. On the other hand, within the high-level GNN, a structure and attribute information fusion module is used to effectively integrate attribute and structural information derived from both the feature encoder (FE) and the graph encoder (GE), thereby enhancing the construction of comprehensive and accurate representations. Our H<sup>2</sup>GnnDTI’s propagation mechanism facilitates the recovery of network properties, such as node degrees and community partitions, and accurately predicts DTIs.

The contributions of this work are summarized as follows:

- 1) H<sup>2</sup>GnnDTI utilizes two levels of GNNs to model drug/protein structures and DTI networks. In the low-level view, it represents drug and protein structures using GNNs: atoms as nodes and chemical bonds or amino acid residues as edges in their respective graphs. This approach integrates atom/residue-level properties and 3D structures synergistically through low-level view GNN (LGNN). In the high-level view, H<sup>2</sup>GnnDTI constructs the DTI graph where drug and protein graphs serve as nodes and interactions as edges, enabling high-level view GNN (HGNN) to learn drug-protein relationships effectively.
- 2) Using an end-to-end training approach, H<sup>2</sup>GnnDTI leverages synergies from both perspectives. The low-level view enriches the high-level view by supplying detailed drug/protein representations, enhancing the accuracy of learned interactive features. Simultaneously, insights gained from drug/protein relationships learned in the high-level view inform refinements in the low-level view, thereby improving the establishment of comprehensive drug/protein representations.
- 3) The structure and attribute information fusion (SAIF) module is used to enhance the elaboration of information from two distinct sources. It integrates sample embeddings from both structural and attribute levels to facilitate consensus representation learning.

We evaluate the performance of our proposed model and compare it with several state-of-the-art models across three datasets. Experimental results consistently demonstrate that H<sup>2</sup>GnnDTI outperforms these models in terms of prediction accuracy, highlighting the strong capability of its hierarchical heterogeneous graph structure in DTI prediction. These

findings suggest that our approach holds promise for advancing the understanding of drug repurposing mechanisms.

## 2 Materials and methods

In this section, we introduce a novel deep-learning model named H<sup>2</sup>GnnDTI for predicting DTIs. Figure 1 provides an overview of the proposed H<sup>2</sup>GnnDTI. Starting with the structures of drugs and targets, H<sup>2</sup>GnnDTI first constructs a hierarchical heterogeneous graph. It uses a low-level encoder to extract structural and chemical features, followed by a high-level encoder to capture interactive features. Finally, DTI predictions are made using a decoder that utilizes the latent features extracted from the graph.

### 2.1 Problem statement

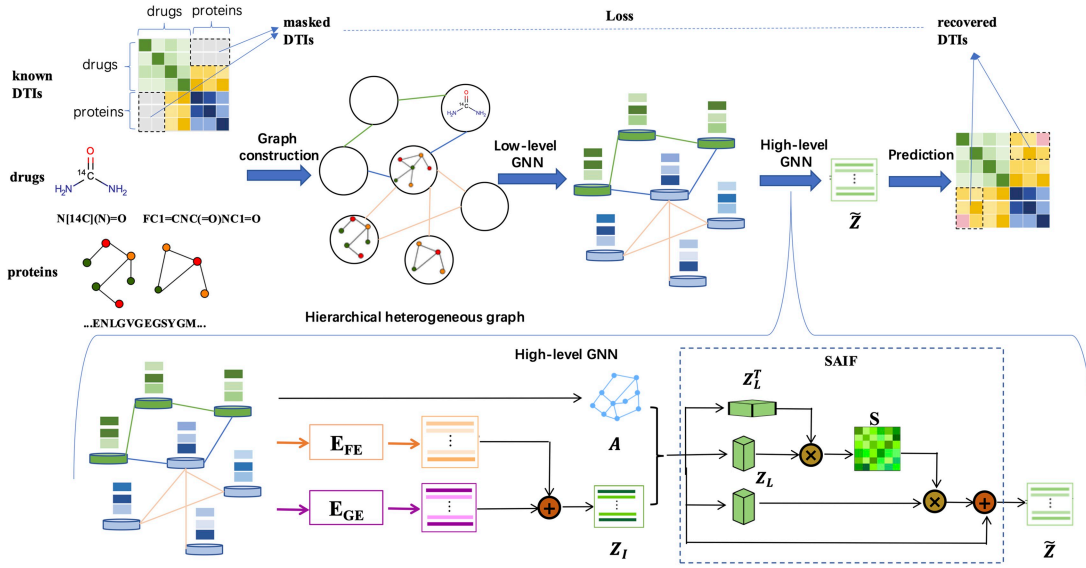
The DTIs can be formulated as an undirected heterogeneous DTI graph  $G = (V, E)$ , where  $V$  represents the set of nodes that contain two disjoint sets of entities from drugs set  $\mathcal{D}$  and protein set  $\mathcal{P}$ ;  $E$  denotes the set of edges representing DTIs. The goal of the DTI prediction is to learn a mapping function  $\tau: E \rightarrow [0, 1]$  from edges to scores, where  $\tau$  is a function with learnable parameter of  $\Theta$ , such that we can determine the probability of drug-target pairs having responses.  $G$  can be represented by node attributes  $X$  and an adjacency matrix  $A$ , where  $A(d, p) = 1$  if drug  $d$  has interaction with protein  $p$  and  $A(d, p) = 0$  otherwise. We denote number of drugs  $|\mathcal{D}| = M$  and number of proteins  $|\mathcal{P}| = N$ .

### 2.2 Hierarchical heterogeneous graph construction

In this work, the hierarchical graph is defined as a high-level heterogeneous graph with each node being a low-level graph. Specifically, the nodes in the high-level graph are drugs and proteins, and the edges are the known DTIs. Each drug or protein node is further represented as a low-level graph with nodes being molecules in each drug or amino acids in each protein with their chemical descriptors. The low-level graphs for drugs and proteins contain their specific structural and chemical information, while the high-level graph involves the binding information.

For the low-level drug graph, each drug is represented as a graph from its atomic sequences. Suppose there are  $m$  atoms in a drug, and each atom is described by  $\mu_1$  kinds of atomic properties, such as the atomic symbol, the atomic degree, the total number of hydrogen atoms, the implicit valence of the atom, whether the bond of the atom is aromatic, etc. The whole list of the characteristics are listed in Supplementary Table S1. The SMILES (Weininger 1988) is a common method for drugs that uses 1D ASCII strings to describe the 3D chemical structure of drugs. In order to take full advantage of the domain knowledge of bio-entities to initialize the node attributes, we use the RDKit technology (an open-source chemistry software) (Bento *et al.* 2020) to convert a drug into a drug molecular graph  $G_d = (V_d, X_d, A_d)$ , where  $V_d$  is the set of drug atom nodes,  $X_d \in \mathbb{R}^{m \times \mu_1}$  is a feature matrix containing the properties of all atoms, and  $A_d$  is an  $m \times m$  adjacency matrix represents the chemical bonds between atoms, i.e. the graph’s edges, in the form of an adjacency matrix.

For the low-level protein graph, each protein is represented as a graph from its amino acid residue sequence. Suppose there are  $n$  residues in a protein, and each residue is described with  $\mu_2$  kinds of physicochemical properties. A protein graph



**Figure 1.** The model structure of H<sup>2</sup>GnnDTI. The inputs include preprocessing drugs and proteins as molecular graphs and protein graphs, respectively, and their known DTIs with masked interactions. With the constructed hierarchical heterogeneous graph, two levels of GNNs (LGNN and HGNN) are sequentially designed to capture hierarchical feature embeddings of drugs and proteins, by propagating information from a molecule in a drug or an amino acid in a protein to other drugs and proteins. The high-level encoder incorporates a SAIF module designed to explicitly integrate representations acquired from both a FE and a GE, facilitating consensus representation learning.

$G_p = (V_p, X_p, A_p)$  is constructed to model the relationship between residues, where  $V_p$  is the set of amino acid residues nodes,  $X_p \in \mathbb{R}^{n \times \mu_2}$  is a feature matrix containing the physicochemical properties of all residues. All the properties used are listed in [Supplementary Table S2](#).  $A_p$  is an  $n \times n$  adjacency matrix representing the connectivity obtained from the protein sequence. The adjacency matrix  $A_p \in \{0, 1\}^{n \times n}$  in the protein graph and protein contact map are exactly equivalent. Contact maps are obtained with atomic level 3D coordinates of proteins. Protein structure information encompasses the angles and distances between different residue pairs, and contact maps are a result of structure prediction. Generally, if the distance between two atoms of residues in Euclidean space is less than a certain threshold ([Wu et al. 2020](#)), they are considered in contact with each other. We can use Pconsc4 (an open-source) ([Michel et al. 2019](#)) to predict contact maps. Next, a protein graph can be obtained by using a threshold of 0.5 to generate the corresponding adjacency matrix. In this graph, residues represent nodes, and the relationships between residues are represented by the adjacency matrix. Each residue node feature is determined by the different functional groups it contains. For a protein sequence with amino acid residues, we use the position-specific scoring matrix as its descriptor introduced by [Jones \(1999\)](#).

For the high-level view, a set of drug graphs and protein graphs can be interconnected within a DTI graph  $G_b = (V_b, X_b, A_b)$ , where  $V_b = \{G_d\}_d \cup \{G_p\}_p$ . The high-level view node attributes is a feature matrix containing the properties of all drug and protein graphs, where  $f$  is an encoder of graph. The  $i$ th row of the feature matrix  $X_b$  represents the representation vector for the  $i$ th drug graph  $G_d$  or the protein graph  $G_p$ . The edges in the high-level graph contains compulsory connections between drugs and proteins, optional connections among drugs and those among proteins, and the adjacent matrix  $A_b$  is an  $(M+N) \times (M+N)$  matrix  $A_b = \begin{pmatrix} S_D & A \\ A^T & S_p \end{pmatrix}$ , where  $A$  denotes the bipartite DTIs among

$M$  drugs and  $N$  proteins,  $S_D$  represents drug similarities and  $S_p$  represents protein similarities. In some cases, especially when the number of drugs and targets are small, drug similarities and protein similarities could be involved in the high-level graph to aggregate information sufficiently. In this case, drug or protein similarities  $S_D$  or  $S_p$  could be calculated from the common number of DTI neighbors. Specifically, if the number of two nodes' common DTI neighbor is greater than a threshold  $\theta$ , the two nodes is connected. A function  $\Psi$  could be used to compute the common neighbor of two nodes ([Wang et al. 2021](#)).

### 2.3 LGNN for learning low-level representations

We use the LGNNs to learn the drug or protein representations. Graph convolutional networks (GCNs) have shown great success for relational data and are suitable for learning graph-structured protein representations. Thus, we propose LGNN based on GCNs. Given the adjacency matrix  $A_d \in \{0, 1\}^{m \times m}$  and the feature matrix  $X_d$  of an arbitrary drug graph  $G_d$ , LGNN adopt a GCN on the molecular graph to learn atom representations which then are summarized into a drug feature vector through a global mean pooling (GEP). Concretely, the  $k$ th layer of the GCN encoder is defined as:

$$H^{(k)} = \text{ReLU}(\tilde{D}^{-\frac{1}{2}}(A_d + I_m)\tilde{D}^{-\frac{1}{2}}H^{(k-1)}W_L^{(k)}), \quad (1)$$

where  $I_m \in \mathbb{R}^{m \times m}$  is the identity matrix,  $\tilde{D} \in \mathbb{R}^{m \times m}$  is the diagonal degree matrix with entries  $D_{aa} = \sum_b (A_d + I_m)_{ab}$ ,  $W_L^{(k)} \in \mathbb{R}^{u_1 \times d_1}$  is a learnable weight matrix for the GCN layer, with  $d_1$  representing the characteristic dimension of each drug molecular after the GCN encoder. ReLU denotes the activation function, and  $H^{(0)} = X_d$ . We then apply the GEP overall molecular graphs, we obtain all drug features that can be compiled into  $X_D \in \mathbb{R}^{|D| \times d_1}$ . To clarify, we use  $x_i \in \mathbb{R}^{1 \times d_1}$  to represent the final representation for the  $i$ th drug graph. Similarly, we can obtain the final representation for the  $j$ th protein graph.

## 2.4 HGNN for learning high-level DTI network

We use the HGNNs to learn DTI network information. HGNN mainly consists of three parts, i.e. a FE, a GE, and a SAIF module (shown in Fig. 1). Based on the features extracted from low-level view, which accounts for structural and chemical information for drugs and proteins, HGNN aims to further capture the interactive features among them.

Formally, we are given the DTI graph  $G_b = (V_b, X_b, A_b)$ , where  $X_b \in R^{(M+N) \times d_1}$  is defined as the feature matrix whose row vector is a final protein representation from LGNN (i.e.  $X_b^{[i,:]} = x_i, i = 1, 2, \dots, M+N$ ). HGNN updates the representation of the high-level graph by a FE and a GE simultaneously.

### 2.4.1 Feature encoder

We first construct a FE with an  $L$ -layer to encode  $X_b$  via

$$Z_{FE}^{(l)} = \Phi(W_{H,F}^{(l)} Z_{FE}^{(l-1)} + b^{(l)}), \quad (2)$$

where  $H^{(0)}$  denotes the final representation  $X_b$  of the low-level view from LGNN, and the activation function

$$\Phi(x) = \text{LeakyReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0. \end{cases}$$

### 2.4.2 Graph encoder

To better make use of both the adjacency information and the attribute information, we design a GE. In the proposed GE, a layer in the encoder is formulated as:

$$Z_{GE}^{(l)} = \sigma_1(D^{-\frac{1}{2}}(A+I)D^{-\frac{1}{2}}Z_{GE}^{(l-1)}W_{H,G}^{(l)}), \quad (3)$$

where  $D \in R^{(M+N) \times (M+N)}$  is the diagonal degree matrix with entries  $D_{ii} = \sum_j (A+I)_{ij}$ ,  $W_{H,G}^{(l)}$  denotes the learnable parameters of the  $l$ th encoder layer, and  $\sigma_1$  is a nonlinear activation function  $\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .

### 2.4.3 SAIF module

Inspired by deep fusion clustering method DFCN (Gong et al. 2022), to sufficiently explore the graph structure and node attributes information extracted by the FE and GE, we take the strategy of the SAIF.

The information integration within our fusion module includes four steps. First, we combine the latent embedding of FE ( $Z_{FE} \in R^{(M+N) \times d_2}$ ) and GE ( $Z_{GE} \in R^{(M+N) \times d_2}$ ) with a linear combination operation:

$$Z_I = \alpha Z_{FE} + (1 - \alpha) Z_{GE}, \quad (4)$$

where  $d_2$  is the latent embedding dimension, and  $\alpha$  is a learnable coefficient which selectively determines the importance of two information sources according to the property of the corresponding dataset. In this work,  $\alpha$  is initialized as 0.5 and then tuned automatically with a gradient decent method.

Then, we process the combined information with a graph convolution-like operation (i.e. message passing operation). With this operation, we enhance the initial fused embedding  $Z_I \in R^{(M+N) \times d_2}$  by considering the local structure within data:

$$Z_L = D^{-\frac{1}{2}}(A+I)D^{-\frac{1}{2}}Z_I, \quad (5)$$

where  $Z_L \in R^{(M+N) \times d_2}$  denotes the local structure enhanced  $Z_I$ .

We further introduce a self-correlated learning mechanism to exploit the nonlocal relationship in the preliminary information fusion space among samples. Specifically, we first calculate the normalized self-correlation matrix  $S \in R^{(M+N) \times (M+N)}$  through:

$$S_{ij} = \frac{e^{(Z_L Z_L^T)_{ij}}}{\sum_{k=1}^{M+N} e^{(Z_L Z_L^T)_{ik}}}. \quad (6)$$

Taking  $S$  as coefficients, we recombine  $Z_L$  by considering the global correlation among samples, and adopt a skip connection to encourage information to pass smoothly within the fusion mechanism:

$$\tilde{Z} = \beta S Z_L + Z_L, \quad (7)$$

where  $\beta$  is a scale parameter. Technically, our fusion mechanism considers the sample correlation in both the perspective of the local and global level. Thus, it has a potential benefit on finely fusing and refining the information from both FE and GE for learning consensus latent representations.

## 2.5 Prediction and loss function

The prediction block consists of the GE decoder, which can reconstruct the adjacency matrix. Inspired by the idea of graph autoencoder, we design a graph decoder that is symmetric to the GE mentioned above:

$$\tilde{Z}^{(b)} = \sigma_1(D^{-\frac{1}{2}}(A+I)D^{-\frac{1}{2}}\tilde{Z}^{(b-1)}W^{(b)}), \quad (8)$$

where  $W^{(b)}$  denote the learnable parameters of the  $b$ th decoder layer, and  $\tilde{Z}^{(0)} = \tilde{Z}$ . The DTI graph is finally reconstructed by

$$\hat{y} = \sigma_2(\tilde{Z}^{(b)} \tilde{Z}^{(b)\top}) \quad (9)$$

from the final layer  $\tilde{Z}^{(b)}$ , and  $\sigma_2$  is the sigmoid activation function  $\sigma_2(x) = \frac{1}{1+e^{-x}}$ .

Let  $\hat{y}$  be the likelihood of masked interactions in the training set. As a binary classification task, we use binary cross entropy loss to train the model:

$$\text{loss} = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}), \quad (10)$$

where  $y$  is the ground truth label for the masked interactions. To infer interactions for new drugs or targets, paired drug and target sequences can be placed in the masked part of the adjacency matrix, forming a large hierarchical heterogeneity graph with the original known data. Then the trained model is used to extract features, and the masked position in the reconstructed adjacency matrix is the predicted result for the new drug or target.

## 3 Experiments

### 3.1 Datasets

We evaluated our H<sup>2</sup>GnnDTI on three real datasets: DrugBank, Davis and KIBA. The DrugBank database

**Table 1.** Statistics of datasets used in our experiments.

| Datasets      | DrugBank | Davis | KIBA   |
|---------------|----------|-------|--------|
| Drug          | 6647     | 68    | 2068   |
| Protein       | 4294     | 379   | 225    |
| Interaction   | 35022    | 25772 | 116350 |
| Positive DTIs | 17511    | 7320  | 22154  |
| Negative DTIs | 17511    | 18452 | 94196  |

(Wishart *et al.* 2006) that was released on 3 January 2020 (version 5.1.5). We removed the drugs which are inorganic compounds, very small molecule compounds or those of which the SMILES string cannot be recognized by RDKit python package (Bento *et al.* 2020). Finally, 6647 drugs, 4294 proteins, and 17 511 positive DTIs are obtained in total. Following the work in Huang *et al.* (2021b) and Wen *et al.* (2017), we randomly sampled positive samples and generated the same number of the negative DTI pair from unknown pairs since no negative DTI pairs are provided. Davis (Davis *et al.* 2011) and KIBA (Tang *et al.* 2014) are two unbalanced benchmark datasets, which provide the binding affinity for every drug–target pair. Following early works (Öztürk *et al.* 2018, Zhao *et al.* 2022), the thresholds 5.0 and 12.1 are set for the Davis and KIBA datasets, respectively, to construct binary classification datasets. The detailed statistics of the three datasets are shown in Table 1.

### 3.2 Comparison methods

We evaluated our H<sup>2</sup>GnnDTI by comparing it with seven state-of-the-art methods as follows.

- **GNN-CPI** (Tsubaki *et al.* 2019): GNN-CPI uses a GNN to encode drugs and uses 1D-CNNs to encode proteins. The latent vectors are then concatenated into a neural network for compound–protein interaction prediction.
- **GNN-PT** (Wang *et al.* 2020): GNN-PT utilizes GNN to extract drug features and uses Transformer and CNN for protein representation. To capture long-range interaction information within proteins, attention mechanisms are incorporated into protein representation modules for CPI modeling.
- **DeepEmbedding-DTI** (Chen *et al.* 2021): DeepEmbedding-DTI utilizes end-to-end representation learning of GNNs with attention mechanism and bidirectional long short-term memory (BiLSTM) to predict DTI.
- **DeepConv-DTI** (Lee *et al.* 2019): DeepConv-DTI captures local residue patterns of targets by convolutions on various lengths of amino acid subsequences. It extracts drug features by using fully connected dense layers on drug fingerprints.
- **TransformerCPI** (Chen *et al.* 2020): TransformerCPI regards drugs and proteins as two kinds of sequences. The features of target sequences are encoded to the decoder. At the same time, the decoder predicts DTIs based on the output of the encoder and the high-levelological features of drugs extracted by GNN.
- **GraphDTA** (Nguyen *et al.* 2019): Compared with DeepDTA, GraphDTA also use the paired drug and target input to calculate its corresponding affinity value, the difference is that the drug is converted into a molecular graph, and then a GNN is used on the graph to learn the representation.

- **MolTrans** (Huang *et al.* 2021a): MolTrans constructs the sub-structural pattern mining algorithm. Then, the interaction map is built from drug sequences and target sequences. Finally, the CNN layer is applied on the interaction map to extract higher-order interactions for predicting DTIs.
- **MCL-DTI** (Qian *et al.* 2023): MCL-DTI combines molecular graph and chemical text data to create comprehensive drug representations. This model adopts a bidirectional multi-head cross attention mechanism and has two independent drug and target decoders.

We follow the same hyper-parameter setting described in the original work for these comparison methods.

### 3.3 Experimental settings

To evaluate the performance of our method, we randomly divided the dataset into training and testing sets with 4:1 ratio. We set up three tasks for DTI prediction, which are commonly encountered in real-world applications. In each setting, different segmentation strategies are used.

- **New-drug (S1)**. Only drug segmentation is used, with interactions between the training drugs and all targets to train the model, which then predicts interactions for test drugs.
- **New-target (S2)**. Only the target segmentation is applied, with interactions between the training targets and all drugs to train the model, which then predicts interactions for test targets.
- **New-drug&New-target (S3)**. Both the drug and target segmentation are used, with the interactions between the training drugs and training targets to train the model, which is then used to predict interactions between the test targets and test drugs.

The H<sup>2</sup>GnnDTI model is trained using the Adam optimizer with a learning rate of  $1e-4$ , and the total epochs is 200. Davis and KIBA are unbalanced benchmark datasets, due to the fact that the positive correlation comes from real interactive datasets, the initial data of DTI is very sparse, we randomly selected negative samples from unrecognized drug target pairs. In the experiment, we set the ratio of positive and negative samples to 1:1. The linear combination parameter  $\alpha$  is set to 0.5, and the skip connection parameter  $\beta$  is set to 0.7. For the three datasets, the graph embedding dimension ( $d_1$ ) of the drug and protein graph is 80, the embedding dimension ( $d_2$ ) of the GE is 20, and the number of layers ( $L$ ) of the GE is 3. Another hyperparameter is the common neighbor threshold,  $\theta$ , for constructing drug similarity graph and protein similarity graph, optionally. Due to the large size of Drugbank dataset, the formed adjacency matrix is too large, we only take known DTIs as the high-level edges to save computational time. For the Davis and KIBA dataset,  $\theta$  is 5 and 3, respectively.

We adopt accuracy (ACC), area under the ROC curve (AUC), and area under the precision-recall curve (AUPR) and F1 as the metrics to measure the performance of models.

### 3.4 Performance evaluation under *de novo* setting

Figure 2 depicts the AUC, AUPR, Precision, Recall, F1, and Acc on DrugBank dataset for the three settings, respectively. H<sup>2</sup>GnnDTI consistently achieves the highest or second highest values for these metrics among all evaluated methods, showing its strong performance for identifying potential

DTIs. This highlights its superior predictive accuracy and robustness compared to other state-of-the-art methods, particularly in challenging scenarios such as new-drug and new-target settings. While the precision by H<sup>2</sup>GnnDTI tends to be the second best one, the highest Recall highlights its ability to accurately identify true positive DTIs, minimizing the risk of missing potential interactions and thus ensuring comprehensive coverage in the prediction process.

Table 2 reports the results for the Davis dataset. As shown in this table, H<sup>2</sup>GnnDTI achieved superior performance over the state-of-the-art baselines on the Davis dataset, with capability to handle realistic situation on drug discovery. Especially, for the challenging New-target setting S2 and New-drug-New target setting S3, where the performance of comparison method tends to decrease dramatically compared to setting S1, our method could perform well and improve the AUC values (and other values) over other methods. The reason is that our model utilizes

hierarchical DTI information and fusion mechanism to dynamically adjust the different features combinations of drugs and proteins, compared to the baselines in which the extracted features of drugs and proteins are often fixed.

Table 3 reports the results for the KIBA dataset. For this dataset, our model outperforms other baseline methods in terms of AUPR and F1 scores significantly for the new-drug and new-target settings, with improvement up to 0.06 for AUPR and 0.09 for F1. However, H<sup>2</sup>GnnDTI tends to produce less competitive AUC and ACC values, possibly due to the large difference in the number of drugs and proteins in the KIBA data, resulting in a highly imbalanced segmentation of the training and testing sets.

### 3.5 Ablation study

To investigate the individual role of each component in the proposed model, we conducted a series of ablation experiments by testing several variants of H<sup>2</sup>GnnDTI:

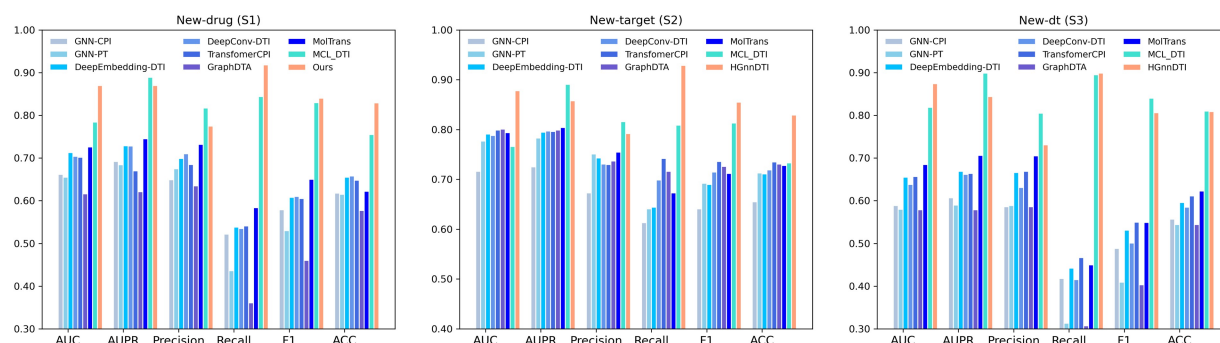


Figure 2. DTI prediction results with other methods under three settings on DrugBank dataset.

Table 2. The four metrics obtained by H<sup>2</sup>GnnDTI and several compared methods in the Davis dataset under three settings, including S1 (New-drug), S2 (New-target), and S3 (New-drug&New-target).<sup>a</sup>

| Settings metrics  | New-drug     |              |              |              | New-target   |              |              |              | New-dt       |              |              |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                   | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          |
| GNN-CPI           | 0.812        | 0.647        | 0.555        | 0.785        | 0.592        | 0.406        | 0.389        | 0.634        | 0.548        | 0.378        | 0.330        | 0.624        |
| GNN-PT            | 0.787        | 0.601        | 0.536        | 0.771        | 0.705        | 0.522        | 0.451        | 0.734        | 0.612        | 0.408        | 0.329        | 0.656        |
| DeepEmbedding-DTI | 0.830        | 0.691        | 0.614        | 0.796        | 0.705        | 0.517        | 0.456        | 0.727        | 0.642        | 0.426        | 0.362        | 0.716        |
| DeepConv-DTI      | 0.845        | 0.692        | 0.631        | 0.800        | 0.673        | 0.455        | 0.460        | 0.701        | 0.640        | 0.420        | 0.381        | 0.640        |
| TransformerCPI    | 0.842        | 0.702        | 0.649        | 0.807        | 0.685        | 0.479        | 0.477        | 0.697        | 0.639        | 0.412        | 0.390        | 0.672        |
| GraphDTA          | 0.802        | 0.641        | 0.514        | 0.789        | 0.690        | 0.451        | 0.342        | 0.695        | 0.605        | 0.360        | 0.206        | 0.675        |
| MolTrans          | 0.791        | 0.621        | 0.696        | 0.774        | 0.661        | 0.452        | 0.406        | 0.711        | 0.642        | 0.419        | 0.391        | 0.648        |
| MCL-DTI           | <u>0.866</u> | <u>0.847</u> | <u>0.768</u> | 0.792        | <u>0.852</u> | <u>0.846</u> | <u>0.780</u> | <b>0.780</b> | <u>0.776</u> | <u>0.725</u> | <u>0.648</u> | 0.697        |
| Ours              | <b>0.883</b> | <b>0.887</b> | <b>0.803</b> | <b>0.811</b> | <b>0.863</b> | <b>0.879</b> | <b>0.790</b> | <u>0.776</u> | <b>0.872</b> | <b>0.884</b> | <b>0.808</b> | <b>0.778</b> |

<sup>a</sup> The best result is marked in bold and the second one is marked with underline.

Table 3. The four metrics obtained by H<sup>2</sup>GnnDTI and several compared methods in the KIBA dataset under three settings, including S1 (New-drug), S2 (New-target), and S3 (New-drug&New-target).<sup>a</sup>

| Settings metrics  | New-drug     |              |              |              | New-target   |              |              |              | New-dt       |              |              |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                   | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          |
| GNN-CPI           | 0.770        | 0.503        | 0.491        | 0.801        | 0.792        | 0.512        | 0.487        | 0.820        | 0.686        | 0.357        | 0.398        | 0.732        |
| GNN-PT            | 0.732        | 0.474        | 0.444        | 0.804        | 0.814        | 0.522        | 0.489        | 0.828        | 0.651        | 0.312        | 0.336        | 0.755        |
| DeepEmbedding-DTI | 0.809        | 0.586        | 0.505        | <b>0.846</b> | 0.792        | 0.507        | 0.484        | 0.808        | 0.706        | 0.379        | 0.369        | 0.780        |
| DeepConv-DTI      | 0.788        | 0.547        | 0.535        | 0.833        | 0.812        | 0.565        | 0.611        | 0.837        | 0.732        | 0.438        | 0.460        | 0.806        |
| TransformerCPI    | 0.799        | 0.555        | 0.528        | 0.830        | 0.822        | 0.581        | 0.522        | 0.833        | 0.717        | 0.416        | 0.404        | 0.795        |
| GraphDTA          | 0.754        | 0.495        | 0.505        | <u>0.836</u> | <u>0.838</u> | 0.619        | 0.556        | <b>0.849</b> | 0.678        | 0.374        | 0.270        | <u>0.821</u> |
| MolTrans          | 0.781        | 0.525        | 0.543        | <u>0.836</u> | 0.817        | 0.573        | 0.604        | <u>0.847</u> | 0.726        | 0.428        | 0.460        | 0.807        |
| MCL-DTI           | <b>0.841</b> | <u>0.770</u> | <u>0.702</u> | 0.781        | <b>0.841</b> | <u>0.767</u> | <u>0.689</u> | <u>0.787</u> | <b>0.912</b> | <b>0.862</b> | <b>0.795</b> | <b>0.839</b> |
| Ours              | 0.798        | <b>0.812</b> | <b>0.743</b> | 0.699        | <b>0.841</b> | <b>0.829</b> | <b>0.774</b> | 0.751        | <u>0.809</u> | <u>0.836</u> | <u>0.792</u> | 0.728        |

<sup>a</sup> The best result is marked in bold and the second one is marked with underline.

- H<sup>2</sup>GnnDTI-L removes the LGNN module and simply uses the original drug/protein sequences. It starts with label encoding to transform each SMILES character and amino acid to the corresponding integer labels, in order to process sequences in batches, the maximum length of the drug is set to 100, parts with a length exceeding 100 will be truncated, strings with a length <100 are padded with 0. The maximum length of an amino acid sequence is 1000.
- H<sup>2</sup>GnnDTI-F only retains the GE in the HGNN, and removes the SAIF fusion mechanism.
- H<sup>2</sup>GnnDTI-W simply replaces the learnable weight  $\alpha$  in Equation (4) by 0.5, i.e. uses the average of the FE and GE as  $Z_I$  in the HGNN.

The performance comparison of H<sup>2</sup>GnnDTI with its three ablation variants on the Davis dataset is shown in Table 4. The experimental results indicate that our proposed model achieves the best performance compared to these variants, demonstrating the individual contributions of our model’s two-level hierarchical view, as well as the advantages of fusion strategies. Among these variant models, the performance degradation of the H<sup>2</sup>GnnDTI-F variant is the most significant, proving that the fusion mechanism is the most critical component of our proposed model. For the S2 task, H<sup>2</sup>GnnDTI-W performs similarly to H<sup>2</sup>GnnDTI, suggesting that the learnable weight  $\alpha$  for the GE and FE in Equation (4) may have a marginal contribution in certain tasks, and that simply averaging could achieve similar results.

### 3.6 Parameter sensitivity analysis

In our work, one hyperparameter is the common neighbor threshold  $\theta$ , which is used to construct the drug–drug edges and target–target edges in the high-level graph. For example,  $\theta = 1$  means that two nodes are connected if they have at least one common neighbor.

We checked the influence of common neighbors under New-drug setting S1 on the Davis dataset. The results of different settings were shown in Fig. 3. From the figure, we could see that a very small  $\theta$  produces lower metrics. Since in DTI network, many different drugs may interact with the same protein as target, a very small  $\theta$  may cause most drugs to be connected, which destroy the similarity information (Wang *et al.* 2021). As the  $\theta$  increases, the DTI graph involves more useful similarity edges, which leads to the accuracy increase, and when  $\theta$  is too big, very few connections between drugs or targets could be detected, which leads to the accuracy decrease. For the Davis and KIBA dataset,  $\theta$  is 5 and 3 respectively.

In addition, the graph embedding dimension ( $d_1$ ) of the drug and protein graph, the embedding dimension ( $d_2$ ) of the GE, and the number of layers ( $L$ ) of the GE will affect the quality of the node embedding representation. These results are available in Supplementary Material.

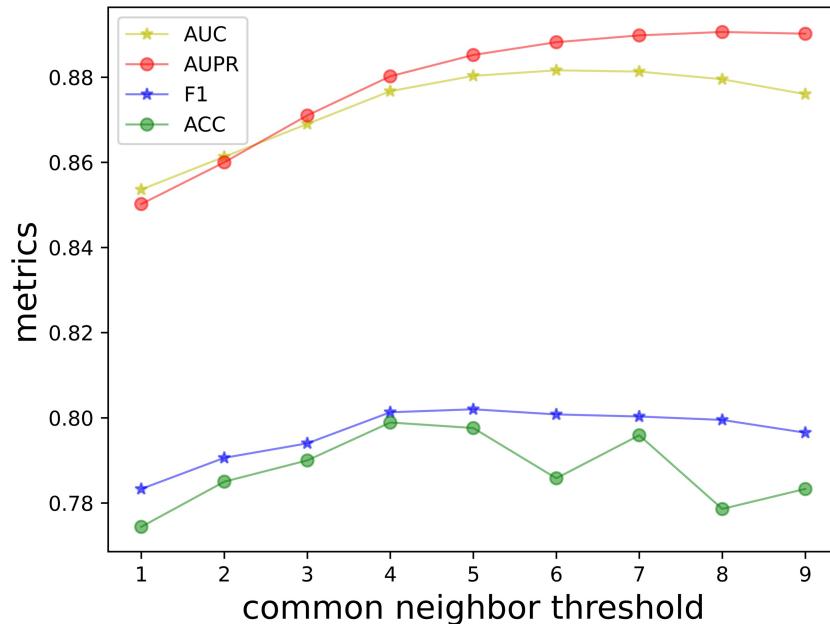
## 4 Conclusion

In this work, we introduce H<sup>2</sup>GnnDTI, a hierarchical graph deep learning approach for predict DTIs, which could

**Table 4.** Ablation study in the Davis dataset under three settings, including S1 (New-drug), S2 (New-target), and S3 (New-drug&New-target).<sup>a</sup>

| Settings metrics        | New-drug     |              |              |              | New-target   |              |              |              | New-dt       |              |              |              |
|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                         | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          | AUC          | AUPR         | F1           | ACC          |
| H <sup>2</sup> GnnDTI   | 0.883        | 0.887        | 0.803        | 0.811        | 0.863        | 0.879        | 0.790        | 0.776        | 0.872        | 0.884        | 0.808        | 0.778        |
| H <sup>2</sup> GnnDTI-L | <u>0.879</u> | <u>0.884</u> | <u>0.798</u> | <u>0.797</u> | 0.857        | <u>0.874</u> | <u>0.787</u> | <u>0.775</u> | <u>0.867</u> | <u>0.881</u> | <u>0.803</u> | <u>0.767</u> |
| H <sup>2</sup> GnnDTI-F | 0.868        | 0.869        | 0.794        | 0.791        | 0.859        | 0.871        | 0.783        | 0.761        | 0.852        | 0.864        | 0.792        | 0.750        |
| H <sup>2</sup> GnnDTI-W | 0.871        | 0.869        | 0.795        | 0.793        | <b>0.865</b> | 0.873        | 0.786        | <b>0.781</b> | 0.855        | 0.864        | 0.799        | 0.756        |

<sup>a</sup> The best result is marked in bold and the second one is marked with underline.



**Figure 3.** Performance evaluation on using different common neighbor threshold for DTI prediction under new-drug setting S1 on the Davis dataset.

simultaneously learn low-level structural and chemical features by low-level GNNs and high-level interactive features by high-level a GNN.

H<sup>2</sup>GnnDTI works like biologists in a hierarchical manner as it contains the two-level view of drug and protein. On one hand, H<sup>2</sup>GnnDTI applies the low-level view when dealing with a drug/protein, which is represented by a drug/protein graph. On the other hand, from the high-level view, protein graphs and their interactions are considered nodes and edges of the DTI graph, respectively. Correspondingly, two GNNs are respectively used to learn from drug/protein graphs in the LGNNs and learn from a DTI graph in the HGNNs. Consequently, a set of graphs are interconnected by edges in a hierarchical graph, to better present the data representation. Moreover, the component SAIF module in the high-level GNN leverages both graph structure and node attributes via a suitable fusion mechanism. In this way, more consensus and discriminative information from both sides is encoded to construct the robust features and improve the generalization capability of the proposed method. Experiments on three benchmark datasets show that H<sup>2</sup>GnnDTI consistently outperforms state-of-the-art baseline methods. In the future, we plan to further improve our method to adapt it to more new drug prediction scenarios.

## Author contributions

Yueying Jing (Resources [equal], Software [lead], Validation [equal], Writing—original draft [equal]) and Dongxue Zhang (Resources [equal], Validation [equal]), Limin Li (Methodology, Writing—review & editing)

## Supplementary data

[Supplementary data](#) are available at *Bioinformatics* online.

Conflict of interest: None declared.

## Funding

This work was supported by National Natural Science Foundation of China projects [92470106 and 12222115].

## Data availability

The preprocessed data used in this work is available at <https://github.com/LiminLi-xjtu/H2GnnDTI/blob/master/dataset.rar>.

## References

- Bai P, Miljković F, John B *et al.* Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat Mach Intell* 2023;5:126–36.
- Bento AP, Hersey A, Félix E *et al.* An open source chemical structure curation pipeline using RDKit. *J Cheminform* 2020;12:51.
- Chen L, Tan X, Wang D *et al.* TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* 2020;36:4406–14.
- Chen W, Chen G, Zhao L *et al.* Predicting drug–target interactions with deep-embedding learning of graphs and sequences. *J Phys Chem A* 2021;125:5633–42.
- Davis MJ, Hunt JP, Herrgard S *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;29:1046–51.
- Gao Z, Jiang C, Zhang J *et al.* Hierarchical graph learning for protein–protein interaction. *Nat Commun* 2023;14:1093.
- Gong L, Tu W, Zhou S *et al.* Deep fusion clustering network with reliable structure preservation. *IEEE Trans Neural Netw Learn Syst* 2022;35:1–12.
- Gu Y, Zheng S, Yin Q *et al.* REDDA: integrating multiple biological relations to heterogeneous graph neural network for drug–disease association prediction. *Comput Biol Med* 2022;150:106127.
- Huang K, Xiao C, Glass LM *et al.* MolTrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics* 2021a;37:830–6.
- Huang K, Xiao C, Glass L *et al.* MolTrans: molecular interaction transformer for drug target interaction prediction. *Oxford Academic* 2021b;37:830–6.
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195–202.
- Kalakoti Y, Yadav S, Sundar D *et al.* TransDTI: transformer-based language models for estimating DTIS and building a drug recommendation workflow. *ACS Omega* 2022;7:2706–17.
- Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations*, 2017.
- Lee I, Keum J, Nam H *et al.* DeepConv-DTI: prediction of drug–target interactions via deep learning with convolution on protein sequences. *PLoS Comput Biol* 2019;15:e1007129.
- Michel M, Menéndez Hurtado D, Elofsson A *et al.* Pconsc4: fast, accurate and hassle-free contact predictions. *Bioinformatics* 2019;35:2677–9.
- Nguyen T, Le H, Quinn T *et al.* GraphDTA: prediction of drug–target binding affinity using graph convolutional networks. *Bioinformatics* 2019;37:1140–7.
- Öztürk H, Özgür A, Ozkirimli E *et al.* DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;34:i821–9.
- Qian Y, Li X, Wu J *et al.* MCL-DTI: using drug multimodal information and bi-directional cross-attention learning method for predicting drug–target interaction. *BMC Bioinformatics* 2023;24:323.
- Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;50:742–54.
- Tang J, Szwajda A, Shakyawar S *et al.* Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model* 2014;54:735–43.
- Tsubaki M, Tomii K, Sese J *et al.* Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* 2019;35:309–18.
- Vaswani A, Shazeer N, Parmar N *et al.* Attention is all you need. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 2017.
- Wang H, Zhou G, Liu S *et al.* Drug–target interaction prediction with graph attention networks. *arXiv*, arXiv:2107.06099, 2021, preprint: not peer reviewed.
- Wang J, Li X, Zhang H. GNN-PT: enhanced prediction of compound–protein interactions by integrating protein transformer. *arXiv*, arXiv:2009.00805, 2020, preprint: not peer reviewed.
- Weininger D. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;28:31–6.
- Wen M, Zhang Z, Niu S *et al.* Deep-learning-based drug–target interaction prediction. *J Proteome Res* 2017;16:1401–9.
- Wishart DS, Knox C, Guo AC *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;34:D668–72.
- Wu Q, Peng Z, Anishchenko I *et al.* Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics* 2020;36:41–8.
- Xu K, Hu W, Leskovec J *et al.* How powerful are graph neural networks? *arXiv*, arXiv:1810.00826, 2018, preprint: not peer reviewed.
- Zhao B-W, You Z-H, Hu L *et al.* A novel method to predict drug–target interactions based on large-scale graph representation learning. *Cancers (Basel)* 2021;13:2111.
- Zhao Q, Zhao H, Zheng KAI *et al.* HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics* 2022;38:655–62. [10.1093/bioinformatics/btab715](https://doi.org/10.1093/bioinformatics/btab715)

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Bioinformatics, 2025, 41, 1–8

<https://doi.org/10.1093/bioinformatics/btaf117>

Original Paper