GENNDTI: Drug-Target Interaction Prediction Using Graph Neural Network Enhanced by Router Nodes

Beiyuan Yang, Yule Liu, Junfeng Wu, Fang Bai, Mingyue Zheng, and Jie Zheng

***Abstract*—Identifying drug-target interactions (DTI) is crucial in drug discovery and repurposing, and in silico techniques for DTI predictions are becoming increasingly important for reducing time and cost. Most interaction-based DTI models rely on the guilt-by-association principle that "similar drugs can interact with similar targets". However, such methods utilize precomputed similarity matrices and cannot dynamically discover intricate correlations. Meanwhile, some methods enrich DTI networks by incorporating additional networks like DDI and PPI networks, enriching biological signals to enhance DTI prediction. While these approaches have achieved promising performance in DTI prediction, such coarse-grained association data do not explain the specific biological mechanisms underlying DTIs. In this work, we propose GENNDTI, which constructs biologically meaningful routers to represent and integrate the salient properties of drugs and targets. Similar drugs or targets connect to more same router nodes, capturing property sharing. In addition, heterogeneous encoders are designed to distinguish different types of interactions, modeling both real and constructed interactions. This strategy enriches graph topology and enhances prediction efficiency as well. We evaluate the proposed method on benchmark datasets, demonstrating comparative performance over existing methods. We specifically analyze router nodes to validate their efficacy in improving predictions and providing biological explanations.**

***Index Terms*—DTI, interpretability, graph enhancement, prior knowledge, graph neural network.**

Manuscript received 18 November 2023; revised 11 March 2024; accepted 8 May 2024. (Beiyuan Yang and Yule Liu are co-first authors.) *(Corresponding author: Jie Zheng.)*

Beiyuan Yang and Yule Liu are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: yangby@shanghaitech.edu.cn; liuyl4@shanghaitech.edu.cn).

Junfeng Wu is with Ainnocence, Shanghai 200082, China (e-mail: jfwu.ai@gmail.com).

Fang Bai is with the Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: baifang@shanghaitech.edu.cn).

Mingyue Zheng is with the Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Beijing 100045, China (e-mail: myzheng@simm.ac.cn).

Jie Zheng is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, and also with the Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai 201210, China (e-mail: zhengjie@shanghaitech.edu.cn).

The proposed method is implemented in Python and the source code can be found at https://github.com/JieZheng-ShanghaiTech/GENNDTI.

Digital Object Identifier 10.1109/JBHI.2024.3402529

I. Introduction

IN DRUG development, identifying drug-target interactions (DTIs) is crucial [1], [2]. DTI aims to locate compounds capable of binding to specific target proteins, aiding in drug virtual screening and repositioning [3]. Traditional methods are often time-consuming and costly, leading to the emergence of data-driven DTI prediction approaches [4], [5], [6]. Docking-based methods, which identify optimal binding sites through molecular simulations, are limited by the precision of 3D structures and slow speed [7], [8], [9], [10]. Machine learning methods use specially designed features to describe drugs and targets. This includes combining structural and evolutionary information [11], constructing kernel functions with molecular descriptors [12], [13] and using techniques like SVM and ensemble learning to focus on important combined features [14]. However, features designed by humans can sometimes introduce biases that make it difficult to accurately capture complex patterns of interaction. With the rise of deep learning and biological data, many studies have applied deep learning models to DTI prediction, mostly using independent feature-based or network-based models [15], [16], [17].

Independent feature-based models focus on exploring the interaction mechanism by employing separate encoders for the drug and target, using inputs like protein sequences and drug SMILES sequences. These models analyze the drug and target features separately [9], [18], [19], [20]. Some common deep learning models used for modeling sequences like CNN [18], LSTM [21], and Transformer [22] have been applied. To overcome the problem that sequence encoders cannot handle topological relationships among atoms in molecules, [23] encodes the drug with graph neural networks (GNNs) to improve prediction accuracy. The study by Wu et al. [24] leverages graph transformer and cross-attention mechanisms to augment the model's capabilities. However, a major limitation of these models is that they find it hard to capture intricate correlations between drugs and targets in DTI prediction [25].

Modeling drug-target interactions as networks is another strategy [26], [27], [28]. These networks are built on the "guilt by association" assumption that similar drugs may act on similar

targets. Hao et al. [29] proposed dual-network integrated logistic matrix factorization to predict DTI by incorporating drug and target profiles. Eslami et al. [30] constructed drug-protein networks using drug-drug and protein-protein similarities and applied graph labeling and deep neural networks to learn complex interaction patterns from embedded graphs. In Shang et al.'s study [31], MEDTI employs multiple similarity networks for compact drug and target feature vectors, enhancing multilayer network representation learning. They include regularization constraints to improve prediction accuracy. Fu et al. [32] build a multi-view heterogeneous network (MVHN) by integrating similarity networks with a biomedical bipartite network. This integration enhances the quality of initial node embeddings, thereby improving prediction efficiency. However, these methods rely on the accuracy and reliability of pre-defined similarity measurements, which can hardly reflect the intricate correlations between drugs and targets and cannot dynamically learn relevant representations to support prediction. In Chu et al.'s work [33], the authors present HGRL-DTA, a model that integrates independent features and known DTA network data as coarse- and fine-level information to predict drug-target affinity. However, it utilizes a fixed message passing strategy to learn two types of information. Some methods enhance drug-target interaction prediction models by incorporating additional drug-drug interaction (DDI) and protein-protein interaction (PPI) networks, which augment the relational data between drugs and targets [34], [35]. Despite these improvements, these models often fail to provide detailed insights into the specific interactions.

To solve these issues, we propose a new strategy that adds special router nodes representing attributes to the existing network. These router nodes reflect the basic properties of drugs and targets. Drugs with similar attributes are connected to the same router drug nodes, and targets with similar attributes are connected to the same router target nodes. When different drugs are linked to the same router drug node, it means these drugs all have that particular characteristic. The same applies to Targets and Router Targets. Therefore, the router nodes and their connections encode fine-grained similarity information between drugs and targets based on their attributes. During graph neural network training, the router nodes provide channels for dynamically propagating semantic messages between similar drugs. This helps the network learn better and make good predictions based on the similarities in features. The study BridgeDPI [36] introduces a similar concept where nodes help to make connections. However, their nodes are virtual ones that link drugs and targets, whereas our router nodes represent real biological features of the drugs or targets. The overall framework of our model is illustrated in Fig. 1. There are four kinds of entities and three types of links. RD stands for router drugs, and RT for router targets. The straight lines between drugs and targets that are known to interact are called real interactions. The lines that connect router drugs to drugs or router targets to targets are called sub-interactions. These sub-interactions help our model better share and understand information, which improves predictions about which drugs might affect which targets. The diagram shows how our router nodes create new paths for information, like linking Drug 1 to Target 3 and Drug 1 to Target 4 through $D_1 \rightarrow RD_1 \rightarrow D_3 \rightarrow T_4$ and $D_1 \rightarrow RD_2 \rightarrow D_4 \rightarrow T_3$.

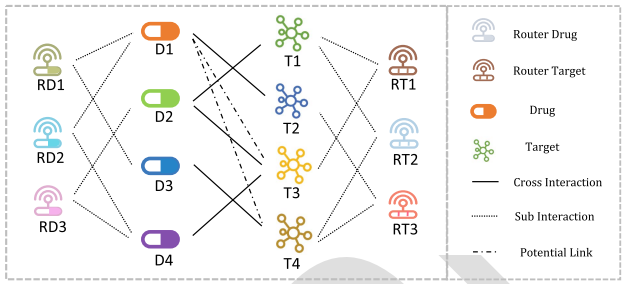


Fig. 1. Overview of our model. The graph contains four types of nodes: drugs, targets, router drugs, and router targets; two types of edges: (1) real interaction between pairs of drugs and targets, (2) sub-interaction between pairs of drugs or targets. The solid line connects the real interactions, and the dotted lines connect the constructed sub-interactions. In the graph, we can see that the introduction of router nodes provides a path from D1 to T3, D1 to T4.

In this work, we propose a novel DTI prediction model named **G**raph **E**nhanced **N**eural **N**etwork for **D**rug **T**arget **I**nteraction Prediction (GENNDTI). This model uses special router nodes that represent the characteristics of drugs and targets, helping to share detailed information effectively. These router nodes are designed to capture and use existing knowledge to highlight similarities between drugs and targets. We use different encoders to process various types of interactions and enhance the drug and target descriptions with molecular fingerprints and amino acid details. The model updates the drug and target information by combining embedding from diverse sources to make accurate predictions. Our approach adds new types of nodes and ways of connecting them in original drug-target interaction networks. By integrating router nodes and various connections, the model clearly distinguishes between different kinds of interactions and provides easy-to-understand explanations.

The main contributions of this paper are listed below:

- We introduce GENNDTI, an innovative model that integrates router nodes representing the attributes of drugs and targets into drug-target interaction networks, enhancing the graph learning process.
- We use distinct graph neural encoders to learn various types of connections within the network.
- We show the interpretability of GENNDTI through case studies, showing how the router nodes visually capture and illustrate semantic similarities.

II. Methods

In this section, we first present the problem definition, then give a detailed description of the GENNDTI architecture. Our proposed framework includes three main parts: (1) A graph enhancement module that involves router nodes and sub-interactions, adding depth and context to the graph structure; (2) a bi-Encoder Representation fusion module that combines the embeddings from different encoders for a comprehensive representation of drugs and targets; (3) an interaction prediction

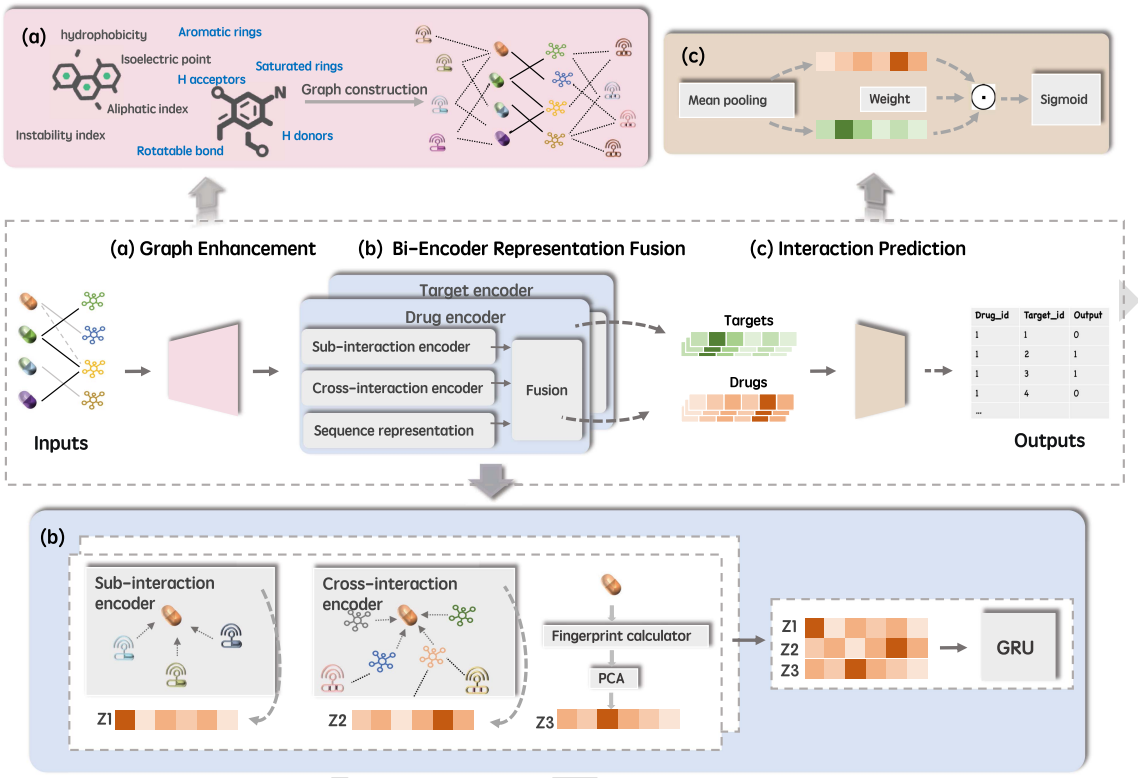


Fig. 2. GENNDTI framework: This framework takes drug-target pairs and existing knowledge about drug proterties and protein characteristics as input to predict how likely different drug-target pairs are to interact. It includes three main parts: (1) A graph enhancement module that uses router nodes and builds additional interactions to improve the network's structure; (2) A Bi-Encoder Representation Fusion module that merges detailed data from various sources, covering both real interactions and sub-interactions; (3) An interaction prediction module that uses the updated node representation to obtain the final drug-target interaction prediction score.

module that uses the updated node representation to obtain the final drug-target interaction prediction score. The framework of GENNDTI is illustrated in Fig. 2.

A. Problem Statement

Let us define $\mathcal{D}$ as the set of all drugs and $\mathcal{T}$ as the set of all targets. Given these definitions, we can represent the dataset and its associated constructs as follows:

- The dataset $\mathcal{X}$ comprises pairs (y_{ij}, r_{ij}), where $y_{ij} = \{(d_i, t_j)|d_i \in \mathcal{D}, t_j \in \mathcal{T}\}$ represents a drug-target pair, and r_{ij} denotes the binary label indicating the presence (1) or absence (0) of an interaction between drug d_i and target t_j.
- We can denote the set of known drug-target interactions (DTIs) as $\mathcal{S} = \{(\mathcal{Y}_{ij}, r_{ij})|(\mathcal{Y}_{ij}, r_{ij}) \in \mathcal{X}, r_{ij} = 1\}$, which includes only those pairs exhibiting interactions.
- The interaction graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ consists of vertices $\mathcal{V} = \mathcal{D} \cup \mathcal{T}$, representing drugs and targets, and edges $\mathcal{E}$, denoting interactions from $\mathcal{S}$ between them.

DTI Prediction Problem: The objective of the Drug-Target Interaction (DTI) problem is to develop a predictive model $\mathbf{F}(\mathcal{Y}_{ij}, \mathcal{G})$ that aims to accurately predict the label r_{ij} between drug i and target j.

B. Graph Enhancement Module

The communication capacity of a graph neural network refers to how well it can exchange and spread information between its nodes. In our model, we enhance this capacity by incorporating additional knowledge into the graph structure. Specifically, we add router nodes between similar node pairs to facilitate message passing. These router nodes act like high-capacity channels that connect potentially interacting drugs and target candidates with shared characteristics. They are designed based on the descriptive chemical features of drugs and targets, enabling them to link related entities and enhance the transmission of relevant information across the network.

Formally, the prior knowledge $\mathcal{K}_d = \{k_{d1}, k_{d2}, k_{d3} \cdots\}$ and $\mathcal{K}_t = \{k_{t1}, k_{t2}, k_{t3} \cdots\}$ are the pre-known properties in drug domain and target domain, respectively. The set of router nodes is denoted by $\mathcal{B}$ and the corresponding edge set is denoted by $\mathcal{E}$. We will add elements to the two sets by the following process.

In the context of drugs, we introduce router nodes represented as b_{di} to integrate prior knowledge denoted by k_{di} forming a set $\mathcal{B}$. Each router node in this collection symbolizes a specific physicochemical characteristic. We then add edges to the edge set $\mathcal{P}$, which represent the drugs that possess these characteristics. The index i is used to count the various prior knowledge features contained in K_d. A similar expansion is applied to

the target domain, with new elements added to both $\mathcal{B}$ and $\mathcal{P}$. Consequently, the expanded graph is structured as follows:

$$\mathcal{J} = \{\mathcal{V} \cup \mathcal{B}, \mathcal{E} \cup \mathcal{P}\} \tag{1}$$

The objective of DTI prediction comes to develop a model $\mathbf{F}(\mathcal{Y}_{ij}, \mathcal{J})$ that can predict the likelihood of interaction, represented as r_{ij} between given drug-target pairs.

In this discussion, we explore the impact of adding extra router nodes and edges on the communication capabilities within drug-target interaction networks. The diameter of a graph, denoted as $D(G)$, represents the longest shortest path between any two nodes in the graph G. Formally, it is the greatest geodesic distance found among any pair of vertices $(u, v) \in V(G)$ and is calculated as the shortest path length $d(u, v)$ between them (2). The diameter essentially measures how far apart the furthest nodes are within the network, considering the shortest path connecting them. A disconnected graph has an infinite diameter, indicating a lack of path connectivity between some node pairs. Conversely, a smaller diameter suggests a more interconnected network, which is advantageous for information flow in Graph Neural Networks (GNNs), enabling more efficient and rapid information propagation [37].

$$D(G) = \max_{u,v \in V(G)} d(u, v) \tag{2}$$

In the graph enhancement module, we add router nodes that connect potentially isolated sections of drug-target interaction (DTI) networks, which can change the network's diameter from infinite to finite. For instance, router nodes that represent common protein functions, like ATP-binding, link separate target entities together. This helps in spreading information between these entities, thereby boosting the network's ability to communicate effectively.

C. Bi-Encoder Representation Fusion Module

GENNDTI aims to improve the communication efficiency of the graph network and the prediction accuracy of the model through special routing nodes and additional connections. The key question we face is how to design the network so that new sub-interaction links can convey information as efficiently and accurately as existing interaction links. Therefore, in this study, we introduce two distinct encoders: one for modeling the original network links (referred to as the cross encoder) and another for the newly added network links (referred to as the inner encoder). This distinction allows us to clearly differentiate between the effects of these two types of interactions, as treating them equally would be inappropriate.

To state how our mechanism works, we will first introduce the general message-passing mechanism [38]. Modern graph neural networks follow a neighborhood aggregation strategy for representation learning on graphs. Specifically, the representation of a node is iteratively updated by aggregating representations of its neighboring nodes. After k iterations, the representation of a node captures topological information within its k-hop neighborhood, which is formulated as follows:

$$m_v^{(k)} = AGGR(e_p^{(k-1)} : p \in \mathcal{N}(v)) \tag{3}$$

$$e_v^{(k)} = \text{COMBINE}(e_v^{(k-1)}, m_v^{(k)}) \tag{4}$$

Here, $m_v^{(k)}$ is the message passing to node v obtained by aggregating the representations of its neighbors, and $e_v^{(k)}$ is the new representation of node v. $\mathcal{N}(v)$ stands for the group of neighboring nodes around node v. In the context of drug-target interactions (DTI), nodes v and p could be drugs, targets, or the special router nodes we've added in our model. We use v and p as general terms for any nodes in the DTI network. The aggregation function $AGGR(\cdot)$ generates messages by aggregating representations of neighboring nodes and the combination function $COMBINE(\cdot)$ fuses the aggregated messages with the node's own representation. Next, we'll explain how we design different message construction methods to reflect various interaction sources.

1) Within Drug or Target Domain: To model the interactions between drugs/targets and their attribute routers, we adopt a message construction scheme following the graph convolutional network. In GCN, neighborhood information is aggregated through Laplacian regularization. The message passed from node p to node v is described as:

$$\mathbf{m}_{vp}^{(k)} = \frac{1}{\sqrt{|\mathcal{N}(p)||\mathcal{N}(v)|}}(\mathbf{W}^{(k)} h_p^{(k)}) \tag{5}$$

Here, $W^{(k-1)}$is a trainable weight matrix. After computing messages from all connected nodes, we add all the messages to form the final message for fusion. The final node v is expressed as:

$$\mathbf{m}_v^{(k)} = \sum_{p \in \mathcal{N}(v) \cap (\mathcal{B} \cup \mathcal{D})} \mathbf{m}_{vp}^{(k-1)} \tag{6}$$

This process allows for increased interaction between similar drugs or targets, thereby enhancing their embeddings.

2) Cross Drug and Target Interaction: This section aims to achieve two core goals by modeling real interactions: (1) making the representation of the interacting drug and target entity pairs similar; (2) making the embedding representations of the associated routers of the interacting drug and target consistent. The core idea behind this goal is that prediction of interactions between drugs and targets will be easier if their representations are highly similar, thus routers should have similar mathematical representations if they are related. We adopted two different methods in this module: Bi-interacion [39] and graph convolutional network (GCN).

In Bi-interacion, we model the similarity between two nodes by taking the element-wise product of the two node embeddings. Specifically, for any two node pairs (v, p) in the drug field, the Bi-interaction message is defined as:

$$\mathbf{m}_{vp}^{(k-1)} = {e_v}^{(k-1)} \odot {e_p}^{(k-1)} \tag{7}$$

Here $\odot$ represents the element-wise product operation, which can encode the similarity between the embedding vectors of nodes v and p. Then, by aggregating the message results between v and all its target neighbors, the final message representation $\mathbf{m}_v^{(k)}$ is obtained.

$$\mathbf{m}_v^{(k)} = \sum_{p \in \mathcal{N}(v) \cap (\mathcal{T} \cup \mathcal{B}_t)} {e_v}^{(k-1)} \odot {e_p}^{(k-1)} \tag{8}$$

where $\mathcal{B}_t$ denotes the set of p's ($p \in \mathcal{T}$) router neighbors. This enables the drug node to obtain information from routers associated with its target nodes. The same method is used for creating messages for target nodes as well.

To improve the accuracy of prediction when data is sparse, we adopt multi-hop propagation and neighbor information aggregation strategies to model the real interactions between drug and targets to make up for the lack of information, thereby obtaining a richer node representation. We obtain the final information of drug node v by aggregating the message passing results between v and all its target neighbors p, whose expression is:

$$\mathbf{m}_v^{(k)} = \sum_{p \in \mathcal{N}(v) \cap (\mathcal{T} \cup \mathcal{B}_t)} \frac{1}{\sqrt{|\mathcal{N}(p)||\mathcal{N}(v)|}} (\mathbf{W}^{(k-1)} e_p^{(k-1)}) \quad (9)$$

where $\mathcal{N}(v)$ and $\mathcal{N}(p)$ represent the neighboring nodes of (v) and (p), respectively.

3) Molecular Representation: To enhance the representation of drug and protein target sequences, we use the RDKit package methods of smiles2morgan and target2aac [40]. The former is used to convert drug information into a numerical representation, while the latter encodes the protein target information into a numerical representation by calculating the composition of amino acid (AA) residues, dipeptides, and tripeptides for a given protein sequence. Then, we use principal component analysis (PCA) to reduce the dimensionality of both numerical representations of the drug and protein target numerical representations from 8420 dimensions to lower dimensions, resulting in a revised representation that preserves important information while eliminating redundancy.

4) Information Fusion: To fuse information effectively, we use the function $COMBINE()$ that transforms data from a 3-dimensional space ($\mathrm{R}^{3\times d}$) into a 1-dimensional space (R^{d}). Previous studies [41] have shown that gated recurrent units (GRU) [42], a recurrent neural network model, are well-suited for consolidating such information. Specifically, messages from sub-interactions and messages from true interactions, along with the node embeddings, are fed as inputs into the GRU. The final output from the GRU gives us the combined node embeddings, which are a rich, integrated representation of the node's information.

$$e = GRU(CONCAT(mv_{\text{node}}, mv_{sub}, mv_{\text{real}}) \quad (10)$$

D. Interaction Prediction

The fused node representations generated in the previous module are fed into a pooling layer to obtain the updated embeddings for drugs or targets, which contain the router information needed for link prediction. We employ a summation operation to obtain the final representations of drug and target nodes.

$$e_d = \sum_{i \in \mathcal{N}(d) \cup d} e_i \quad (11)$$

$$e_t = \sum_{j \in \mathcal{N}(t) \cup t} e_j \quad (12)$$

TABLE I
STATISTICS OF TWO DTI DATASETS

	Drugs	Target	Interaction	Positive pairs	Negative Pairs
Davis	68	379	25772	7429	18343
KIBA	2068	229	117657	22566	95091

The probability of interaction between drug e_d and target e_t is calculated by the following formula:

$$\hat{r}_{i,j} = \phi(f(e_d, e_t)) \quad (13)$$

where f is the inner product function and ϕ is the sigmoid function that limits the score to the interval between 0 and 1. We set 0.5 as a threshold to convert the output values into binary labels indicating whether there is an interaction between the candidate drug target pairs.

E. Optimization Objective and Loss Function

The optimization objective of GENNDTI consists of two parts, a base loss function and an L_2 regularization term. The base loss function uses binary cross-entropy to quantify the difference between the true labels and the predicted labels, which can be expressed as:

$$L(\hat{r}_{i,j}(\theta), r_{i,j}) = -r_{i,j} \cdot \log(\hat{r}_{i,j}(\theta)) + (1 - r_{i,j}) \cdot \log(1 - \hat{r}_{i,j}(\theta)) \quad (14)$$

where $r_{i,j}$ is the true label for sample i and sample j, and $\hat{r}_{i,j}(\theta)$ is the predicted label under parameters θ. To prevent overfitting, an L_2 regularization term is introduced, expressed specifically as:

$$R = \lambda(\|\theta\|^2) \quad (15)$$

The final optimization goal of GENNDTI is represented as:

$$L(\theta) = \frac{1}{N} \sum_{n=1}^{N} L(\hat{r}_{i,j}(\theta), r_{i,j}) + R \quad (16)$$

$$\theta^* = \arg\min_{\theta} L(\theta) \quad (17)$$

N is the total number of samples, θ represents all the parameters, and θ^* are the final parameters after optimization.

III. Experimental Results

In this section, we first describe the experimental setup. Then we show the performance of GENNDTI by comparing it with the state-of-the-art models, followed by a comparative study and an ablation study to understand the effectiveness of each component in GENNDTI. Finally, we analyse how the routers impact DTI prediction.

A. Data Preparation

1) Datasets of DTI Pairs: In this study, we use the benchmark datasets Davis [43] and KIBA [44] to evaluate the model performance. The statistic of the two datasets are given in Table I.

Davis: Davis contains binding affinities between 68 drugs and 379 proteins, constituting 25,772 DTI pairs. It includes

TABLE II
DESCRIPTORS OF THE ATTRIBUTES OF DRUGS AND TARGETS (# MEANS THE NUMBER OF)

Drug Descriptors	Target Descriptors
# Aromatic Carbocycles	Aliphatic index of peptide
# Aromatic Heterocycles	Potential peptide interaction index
# Aromatic Rings	Hydrophobicity index
# H Acceptors	instability index
# H Donors	Isoelectric point
# Heteroatoms	Quality difference of modified peptides
# Rotatable Bonds	
# Saturated Carbocycles	
# Saturated Heterocycles	
# Saturated Rings	

the results of selectivity assays for the kinase protein family and their inhibitors, along with their dissociation constant (K_d) values. We transform K_d into $-\log_{10}(\frac{K_d}{1e9})$ for data splitting in logspace. Following the experimental setting of [45], we divide the Davis dataset by the threshold 5.0 to construct a binary classification database with a connectivity of 28.8%. The instances that surpass the value of 5.0 are considered positive samples, while those below 5.0 are regarded as negative samples.

KIBA: KIBA contains binding affinities between 2,068 drugs and 229 proteins, together constituting 117,657 drug-target interaction pairs. The KIBA dataset encompasses selectivity assays conducted on kinase proteins and their corresponding inhibitors. The KIBA scores are calculated from experimental data, specifically K_i, K_d, and IC_{50} values, gathered from trusted sources. We used a threshold of 12.1 following [12] for data partitioning of the processed KIBA dataset, which formed a database with a connectivity of 4.76%. The instances that surpass the value of 12.1 are considered positive samples, while those below 12.1 are regarded as negative samples.

The connectivity is defined as

$$\text{connectivity} = \frac{\text{existing connections}}{\text{number of drugs} \times \text{number of targets}} \quad (18)$$

2) Prior Knowledge of Molecular Attribtues: We use prior information about the characteristics of drugs and targets to create "router nodes," each representing a specific characteristic. For the drugs, we select a set of descriptors from the RDKit.Chem.Descriptors module [46]. This module is a Python package that provides 208 descriptors, mainly consisting of physicochemical properties and fractions of substructures in the drugs. For the targets, we selected some protein descriptors from the Peptides package [47], which is a Python package that contains physicochemical properties, indices, and descriptors for amino acid sequences. The descriptors that we have chosen are shown in Table II. When picking router nodes, we focus on attributes that can be turned into whole numbers or grouped into specific categories. This strategy simplifies the ways to enhance the probability of establishing connections between diverse drugs and targets to the same router node.

B. Experimental Setting

1) Metrics: We evaluate the model using commonly used performance metrics for binary classification, including AUC (Area Under the Curve), AUPR (Area Under the Precision-Recall Curve), accuracy, precision, and recall.

2) Implementation Details: We implemented our model with Pytorch 1.6.0 and PyTorch Geometric 1.4.3, and conducted the training and testing phases on two NVIDIA 2080 Ti GPUs. We obtained the datasets from the Therapeutics Data Commons (TDC) [48]. We divided each dataset into training, validation, and test sets in a 7:1:2 ratio, respectively. The validation set facilitated the determination of hyperparameter configurations, whereas the test set served for model performance evaluation. To ensure the reliability of our results, we conducted 5 independent trials for each experiment. We considered the mean score across these 5 trials as the final result. The hyperparameters selected for our model are detailed in Table V.

3) Baselines: We compared our model with several recently proposed baseline methods for DTI prediction, which include GNN-CPI [49], GNN-PT [50], DeepEmbedding-DTI [51], GraphDTA [23] (Which was originally designed for the regression problem of predicting binding affinity, but can be converted to a binary classifier by adding a sigmoid function to the output layer), DeepConv-DTI [16], TransformerCPI [22], MolTrans [2], GCN [52], BridgeDPI [36]and HGRL-DTA [33]. We adopt the same data partition as [45]. For BridgeDPI, We reproduce the article with the parameters the original paper provides [36].

C. The Prediction Ability

The experimental results in Tables III and IV indicate that our proposed GENNDTI model achieves competitive performance in the majority of cases, which demonstrates the effectiveness of our model.

We achieved noteworthy results on Davis, where our model has improved precision, recall, AUC, and AUPR by 1.7%, 2.6%, 0.8%, and 0.3% compared to the best baseline model. Simultaneously, our model also achieved competitive results on KIBA. As previously mentioned, the models we compared are of two types: those based on independent features (the first seven models) and those based on interactions (the last four). In From the data in Tables III and IV, we see that models focusing on interactions worked better on the dense Davis dataset. However, for the sparser KIBA dataset, there wasn't a big difference in how the two types of models performed. Interaction-based methods leverage message passing between neighboring nodes, while independent feature-based models predict based on separate drug-target pairs. This means interaction-based models do well when the network of connections is strong and close. The Davis dataset, which is well-connected, allows for a lot of information sharing, helping our model perform well. However, the KIBA dataset stayed sparse, even after attempts to enhance it, leading to less message sharing and only a slight improvement in predictions.

Among all baseline models compared, GraphDTA, HGRL-DTA and MolTrans showed better performance. In particular, GraphDTA and HGRL-DTA adopt graph neural networks (GNN) to model molecular graphs, which demonstrates the effectiveness of using GNN to characterize molecular structures to improve drug target affinity prediction. The GENNDTI model increases the density of the network by incorporating prior knowledge and promoting information exchange between

TABLE III
PERFORMANCE COMPARISON OF GENNDTI WITH BASELINES IN AUC AND AUPR ON DAVIS (STD)

	Accuracy(Std)	Precision(Std)	Recall(Std)	AUC(Std)	AUPR(Std)
GNN-CPI	0.819 (0.001)	0.731 (0.002)	0.570 (0.002)	0.863 (0.001)	0.745 (0.002)
GNN-PT	0.827 (0.001)	0.693 (0.020)	0.706 (0.021)	0.882 (0.007)	0.774 (0.010)
DeepEmbedding-DTI	0.836 (0.008)	0.760 (0.017)	0.618 (0.024)	0.878 (0.011)	0.775 (0.020)
DeepConv-DTI	0.830 (0.001)	0.750 (0.002)	0.698 (0.001)	0.867 (0.001)	0.777 (0.001)
TransformerCPI	0.822 (0.001)	0.688 (0.003)	0.688 (0.003)	0.877 (0.001)	0.767 (0.001)
MolTrans	0.842 (0.000)	0.782 (0.003)	0.617 (0.004)	0.900 (0.001)	0.784 (0.002)
GraphDTA	0.817 (0.001)	0.743 (0.014)	0.530 (0.017)	0.859 (0.004)	0.743 (0.007)
GCN	0.925 (0.004)	0.889 (0.003)	0.872 (0.004)	0.887 (0.002)	0.950 (0.001)
HGRL-DTA	**0.939 (0.003)**	0.809 (0.005)	0.679 (0.006)	0.832 (0.003)	0.757 (0.001)
BridgeDPI	0.931 (0.003)	0.946 (0.005)	0.931 (0.003)	0.885 (0.014)	0.987 (0.002)
GENNDTI-BI	0.935 (0.001)	0.952 (0.002)	**0.964 (0.002)**	**0.908 (0.003)**	**0.990 (0.001)**
GENNDTI-GCN	0.933 (0.001)	**0.954 (0.004)**	0.963 (0.001)	0.892 (0.005)	0.988 (0.001)

The baseline results are from[45]. Bold: optimal performance, underline:sub-optimal.
GENNDTI-BI represents (Bi-interaction,GCN) and GENNDTI-GCN represents (GCN,GCN) in the comparative study.

TABLE IV
PERFORMANCE COMPARISON OF GENNDTI WITH BASELINES IN AUC AND AUPR ON KIBA (STD)

	Accuracy(Std)	Precision(Std)	Recall(Std)	AUC(Std)	AUPR(Std)
GNN-CPI	0.867 (0.002)	0.727 (0.002)	0.477 (0.007)	0.864 (0.005)	0.673 (0.005)
GNN-PT	0.876 (0.005)	0.691 (0.006)	0.647 (0.007)	0.901 (0.002)	0.741 (0.005)
DeepEmbedding-DTI	0.878 (0.002)	0.741 (0.005)	0.556 (0.016)	0.889 (0.003)	0.727 (0.006)
DeepConv-DTI	0.878 (0.001)	0.708 (0.002)	0.636 (0.003)	0.898 (0.001)	0.703 (0.001)
TransformerCPI	0.870 (0.001)	0.669 (0.003)	0.631 (0.003)	0.888 (0.001)	0.708 (0.001)
MolTrans	0.881 (0.001)	0.710 (0.003)	0.645 (0.003)	0.905 (0.001)	0.708 (0.003)
GraphDTA	0.889 (0.001)	0.775 (0.020)	0.594 (0.032)	**0.914 (0.001)**	0.776 (0.007)
GCN	0.873 (0.003)	0.721 (0.013)	0.603 (0.012)	0.874 (0.001)	0.705 (0.006)
HGRL-DTA	**0.904 (0.003)**	**0.783 (0.005)**	0.670 (0.002)	0.862 (0.002)	0.769 (0.003)
BridgeDPI	0.876 (0.003)	0.743 (0.003)	0.668 (0.004)	0.903 (0.002)	0.762 (0.004)
GENNDTI-BI	0.873 (0.002)	0.741 (0.014)	0.662 (0.008)	0.902 (0.002)	0.760 (0.005)
GENNDTI-GCN	0.874 (0.001)	0.746 (0.018)	**0.671 (0.005)**	0.908 (0.001)	**0.778 (0.003)**

The baseline results are from[45]. Bold: optimal performance, underline:sub-optimal.
GENNDTI-BI represents (Bi-interaction,GCN) and GENNDTI-GCN represents (GCN,GCN) in the comparative study.

similar molecules. This enhances the model's ability to understand molecular properties by strengthening the message passing mechanism, thereby improving the accuracy of DTI predictions.

D. Model Analysis

In this section, we will look into four key questions: (1) Should we use different methods to model the real interaction and sub-interactions (2) Do the sub-interactions positively affect the final predictions? (3) How important is each component of the model? (4) What's the best way to combine various types of information?

1) Comparative Study of Different Combinations of Interactions: On the choice of two interaction modes, we referred to similar work [41]. Specifically, we use (real-interaction, sub-interaction) pairs to represent different combined versions of model selection. For example, (Bi-interaction, GCN) indicates that Bi-interaction and GCN are used to model real interactions and sub-interactions, respectively.

We use Bi-interaction and GCN to simulate real interactions and MLP [53] and GCN to simulate sub-interactions, respectively. By pairing the results of these two interaction modelings, we tested on the Davis and KIBA datasets, and the experimental results are shown in Fig. 3.

Fig. 3 shows that the Graph Convolutional Network (GCN) is better than the Multi-Layer Perceptron (MLP) at modeling sub-interactions. This finding indicates that using prior knowledge to build sub-interactions is more effectively enhanced by first merging information and then combining this merged information with other relevant data. Therefore, GCN is more appropriate for these tasks than MLP.

For real interaction, the effectiveness varies with the dataset's connectivity. For densely connected datasets like Davis, the Bi-interaction method outperforms GCN, suggesting it's better

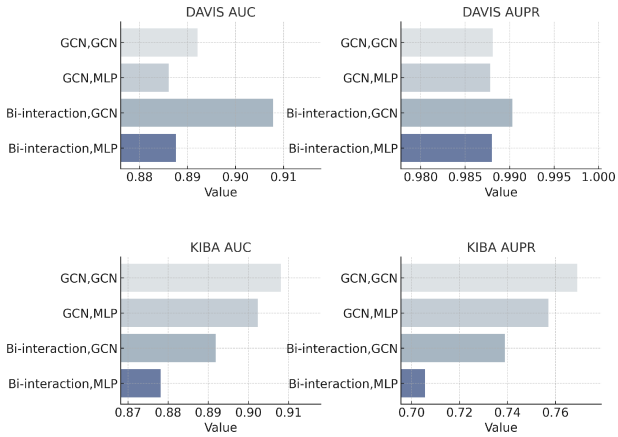


Fig. 3. The comparison of using different combinations of interaction and neural network models.

TABLE V
HYPERPARAMETERS OF GENNDTI

Hyperparameter	Value
dimension	64
hidden_layers	256
L2 weight	1e-5
learning rate	6e4
batch size	128
epoch	100
early stop patience	15
optimizer	Adam

TABLE VI
THE COMPARISON WHEN INCLUDING DIFFERENT ROUTER NODES(STD)

	Davis		KIBA	
	AUC	AUPR	AUC	AUPR
w/o routers	0.877	0.987	0.872	0.699
router Drug only	0.891	0.988	0.904	0.762
router Target only	0.905	0.990	0.880	0.715
both routers	0.908	0.990	0.908	0.778

suited for dense networks. However, for sparser networks like KIBA, GCN is more effective than Bi-interaction. This difference implies that sparser networks benefit from more rounds of information gathering to improve outcomes, whereas dense networks might need only one. Too many rounds of combining information in a dense network could result in over-smoothing.

2) Ablation Study of Routers: To evaluate the impact of including routers and sub-interactions in our model, we carried out an ablation study. This study involved using different combinations to encode the real interactions and the sub-interactions within the model. Specifically, for the Davis dataset, we employed the (Bi-interaction and GCN) combination to encode real interactions and sub-interactions and for the KIBA dataset, we used "GCN and GCN," in line with our prior discussions.

We conducted experiments on two datasets under the following four settings: (1) without using any routers, (2) using only drug routers, (3) using only target routers, and (4) using both drug and target routers. The results are in Table VI.

The results clearly indicate that adding router nodes makes the model work better. Out of all the setups we tested, the one with both drug and target routers gave the best results. This shows that the model really benefits from having a complete set of routers. On the other side, not using any routers at all gave the worst outcomes. Also, just including one type of router, either for drugs or targets, still helped improve the model's performance, but the extent of improvement varied based on the setup.

The Davis dataset has more target router nodes than drug ones, enhancing the model more with target information. Conversely, the KIBA dataset has more drug router nodes, so drug information boosts the model significantly. Overall, the model performs best when it includes both types of information.

TABLE VII
ABLATION STUDY OF DIFFERENT MODULES(STD)

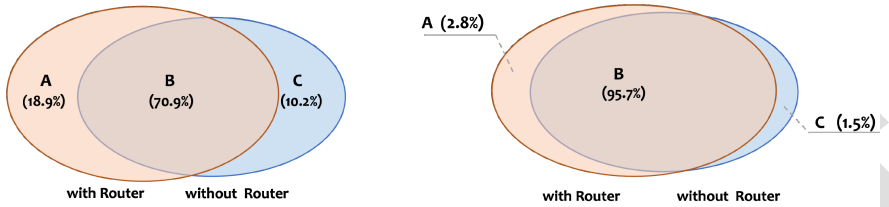
	Davis		KIBA	
	AUC(Std)	Accuracy(Std)	AUC(Std)	Accuracy(Std)
w/o cross	0.839	0.925	0.886	0.866
w/o sub	0.863	0.923	0.889	0.868
w/o fingerprint	0.871	0.929	0.903	0.869
whole model	0.908	0.935	0.908	0.874

TABLE VIII
ABLATION STUDY OF FUSION METHODS(STD)

	Davis		KIBA	
	AUC(Std)	Accuracy(Std)	AUC(Std)	Accuracy(Std)
SUM	0.838	0.926	0.902	0.871
MLP	0.704	0.914	0.831	0.832
GRU	0.908	0.935	0.908	0.874

3) Ablation Study of Different Modules: To check how effective the three modules in the Drug encoder and Target encoder are, we run tests removing each module one by one. We look at how the model performed without the cross-interaction module, without the sub-interaction module, without the fingerprint module, and compared these with the performance of the full model, focusing on the AUC and Accuracy metrics. Experimental results show that the complete model, with all modules included, works better than any version with a module removed. This means each module adds value to the model. Removing the cross-interaction module resulted in the most significant performance decline, showing its vital importance in understanding the interactions between known drugs and targets, which is key for accurately predicting drug-target interactions (DTI). The results are in Table VII.

4) Ablation Study of Fusion Method: The fusion module in GENNDTI aggregates messages from drug encoders and target encoders to obtain fused node representations of drugs and targets. We evaluated three information fusion methods: SUM [54], MLP, and GRU [55]. SUM uses element-wise addition across the vectors to fuse their information. MLP employs a Multilayer Perceptron to learn nonlinear combinations of vectors, which involves processing through linear transformations and nonlinear activation layers to get a node representation that integrates various information features. Lastly, GRU utilizes a gated mechanism to adaptively fuse information from different sources. The experimental results, as shown in Table VIII, indicate that the GRU is the most effective fusion method of the three types of information. It surpasses other methods by



(a) True positive node pairs in KIBA (b) True negative node pairs in KIBA

Fig. 4. The Venn diagrams show the proportion of predicted node pairs under the With Router condition and without Router condition. Figure a) shows the distribution of correctly predicted interacting node pairs (true positives) under both settings. Figure b) depicts the distribution of correctly predicted non-interacting node pairs (true negatives) under the two scenarios.

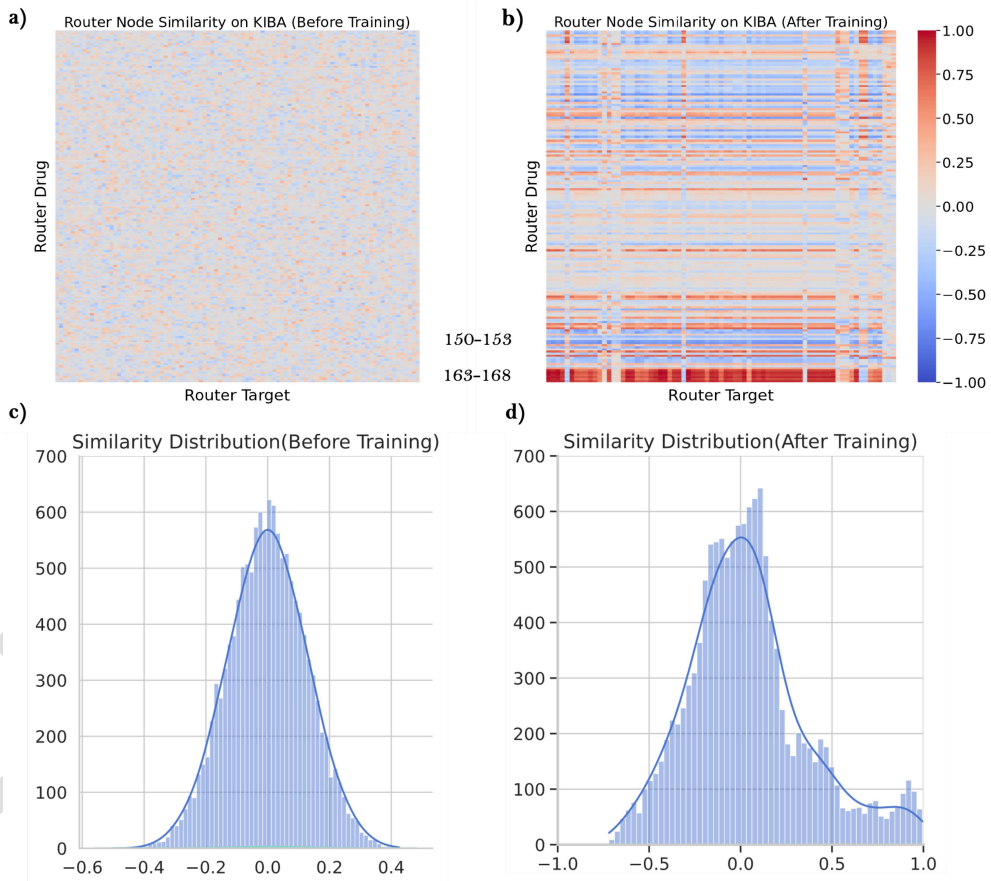


Fig. 5. Case study results about the similarity between property embeddings of router drugs and targets. (a) Heatmap that visualises the correlations (similarities) between drug and target properties as before training, (b) Heatmap of the correlations and target properties after training. (c) The distribution of the property similarities before training. (d) The distribution of the property similarities after training.

providing more flexibility compared to SUM's simple addition and more efficiency than MLP's static layers.

E. The Impact of Routers

In this section, we aim to explore the function of router nodes in the model's predictions and provide their biological interpretation.

1) Improvements to Graph Topology: We conducted a visual analysis of the prediction results on the test set of the KIBA dataset under two different scenarios, i.e. with and without router nodes. The results are presented in Fig. 4. In Fig. 4(a), the prediction outcomes for true positive node pairs are shown, and in Fig. 4(b), the prediction outcomes for true negative node pairs are illustrated in Venn diagrams. In the context of the analysis, A and C represent the sets of node pairs that can be predicted only with router nodes and only without router nodes, respectively. B represents the sets of node pairs that can be predicted accurately under both settings. We can see that the introduction of routers leads to an increase in the count of correctly predicted node pairs.

We isolated node pairs accurately predicted only by the model incorporating routers and examined the degrees of these nodes in

the bipartite graph derived from the original dataset. The degrees were lower compared to the average degree of the entire network. This observation indicates router nodes mitigate insufficient connectivity for certain challenging nodes, thereby enhancing the graph's learning capacity. For instance, in the KIBA dataset graph, the average target node degree is 108.035. However, for target nodes from the original graph solely predictable with router augmentation, the average degree is just 16.57. This discrepancy highlights the presence of relatively disconnected nodes in the graph. Augmenting their connections via router nodes improves predictive performance by alleviating such insufficient connectivity. As stated in [56], graph sparsity impedes representational power. Correspondingly, our method enhances expressivity by increasing graph density through introduced router nodes, aligning with referenced conclusions.

2) Interpretability of Router Nodes: To illustrate the representation learned by router nodes, we extract router embeddings before and after training. We generated heatmaps to visualize the cosine similarities between router drugs and router target embeddings, as well as the distribution of these similarities, and the final results are shown in Fig. 5.

Our approach incorporates router nodes based on prior knowledge to serve as intermediaries between drug and target entities. This enables elucidating the relevance of specific attributes in predicting drug-target interactions. We conducted a case study on the KIBA dataset to validate the efficacy of router nodes. In Fig. 5, target router embeddings are plotted horizontally while drug routers are vertically in the heatmaps. Initially, the router embeddings are randomly initialized before training, yielding correlations centered around 0 as depicted in Fig. 5(c). This indicates the absence of discernible patterns between the untrained routers. However, after training, the correlation distribution undergoes notable changes as shown in Fig. 5(d). Certain router pairs exhibit highly positive correlations approaching 1, implying strong relevance. Conversely, some pairs display highly negative correlations near -0.7, indicating opposing traits. Nonetheless, most router pairs lack significant correlations. These observations demonstrate our approach successfully learns salient property relationships of router nodes. By training, the router correlations become more reflective of intrinsic drug-target interaction patterns.

As shown in Fig. 5(b), router drugs within the 163–168 range exhibit strong correlation with numerous target routers. These routers represent high quantities of saturated heterocycles in the molecules. This highlights the salient role of specific heterocycles in drug discovery, as their presence or absence remarkably affects interaction with certain targets. Prior studies have demonstrated hydrogen-bond acceptors of heteroatoms can bind proteins [57] and marketed drugs with high affinity often contain ring structures [57], corroborating our observation. In contrast, drugs with 150–153 router embeddings show strongly negative correlations, potentially hindering interactions. These routers represent molecules with many ($\geq$10) rotatable bonds, aligning with Lipinski's Rule of Five that excessive flexibility from rotatable bonds may reduce protein binding [58].

For knowledge validation, routers and sub-interactions can be constructed from prior domain expertise to examine the influence of specific molecular subcomponents on the interaction mechanism. Hence, this analysis signifies GENNDTI's potential as an invaluable tool for knowledge discovery to drug-target interactions.

IV. Conclusion and Discussion

DTI prediction is critical for drug discovery and repositioning. Most interaction-based models rely on the guilt-by-association principle. However, they cannot dynamically extract complex correlations or reflect specific causal factors underlying interactions. In our paper, we introduce a new model called GENNDTI, which introduces router nodes based on biological knowledge to construct paths for message passing and uses diverse encoders to distinguish interaction types. These routers act as interpretable passageways that propagate informative signals between drug and target nodes. By learning on the enhanced graph, our approach not only accurately predicts Drug-Target interactions (DTIs) but also provides insights into the underlying mechanisms. We demonstrate GENNDTI's superiority over existing approaches on several benchmark datasets. We also evaluated the strengths and weaknesses of different types of methods on diverse connectivity datasets. Furthermore, we validated the contribution of router nodes in enhancing model performance and biological interpretability. In future work, we plan to incorporate more biological information using hypergraph neural networks or other techniques to to further explore the DTI response mechanism.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable suggestions.

References

[1] C. V. Theodoris et al., "Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease," *Science*, vol. 371, no. 6530, 2021, Art. no. eabd0724.

[2] K. Huang, C. Xiao, L. M. Glass, and J. Sun, "MolTrans: Molecular interaction transformer for drug–target interaction prediction," *Bioinformatics*, vol. 37, no. 6, pp. 830–836, 2021.

[3] D.-L. Ma, D. S.-H. Chan, and C.-H. Leung, "Drug repositioning by structure-based virtual screening," *Chem. Soc. Rev.*, vol. 42, no. 5, pp. 2130–2141, 2013.

[4] M. Bagherian, E. Sabeti, K. Wang, M. A. Sartor, Z. Nikolovska-Coleska, and K. Najarian, "Machine learning approaches and databases for prediction of drug–target interaction: A survey paper," *Brief. Bioinf.*, vol. 22, no. 1, pp. 247–269, 2021.

[5] R. Chen, X. Liu, S. Jin, J. Lin, and J. Liu, "Machine learning for drug-target interaction prediction," *Molecules*, vol. 23, no. 9, 2018, Art. no. 2208.

[6] X. Ru, X. Ye, T. Sakurai, Q. Zou, L. Xu, and C. Lin, "Current status and future prospects of drug–target interaction prediction," *Brief. Funct. Genomic.*, vol. 20, no. 5, pp. 312–322, 2021.

[7] N. Nagamine and Y. Sakakibara, "Statistical prediction of protein–chemical interactions based on chemical structure and mass spectrometry data," *Bioinformatics*, vol. 23, no. 15, pp. 2004–2012, 2007.

[8] Z. He et al., "Predicting drug-target interaction networks based on functional groups and biological features," *PLoS One*, vol. 5, no. 3, 2010, Art. no. e9603.

[9] Y. Wang and J. Zeng, "Predicting drug-target interactions using restricted Boltzmann machines," *Bioinformatics*, vol. 29, no. 13, pp. i126–i134, 2013.

[10] Q. Yuan, J. Gao, D. Wu, S. Zhang, H. Mamitsuka, and S. Zhu, "DrugE-Rank: Improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank," *Bioinformatics*, vol. 32, no. 12, pp. i18–i27, 2016.

[11] F. Rayhan et al., "iDTI-ESBoost: Identification of drug target interaction using evolutionary and structural features with boosting," *Sci. Rep.*, vol. 7, no. 1, pp. 1–18, 2017.

[12] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: A read-across approach for predicting drug–target binding affinities using gradient boosting machines," *J. Cheminformatics*, vol. 9, no. 1, pp. 1–14, 2017.

[13] A. Cichonska et al., "Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors," *PLoS Comput. Biol.*, vol. 13, no. 8, 2017, Art. no. e1005678.

[14] Y. Gong, B. Liao, P. Wang, and Q. Zou, "DrugHybrid_BS: Using hybrid feature combined with bagging-SVM to predict potentially druggable proteins," *Front. Pharmacol.*, vol. 12, 2021, Art. no. 771808.

[15] M. Wen et al., "Deep-learning-based drug–target interaction prediction," *J. Proteome Res.*, vol. 16, no. 4, pp. 1401–1409, 2017.

[16] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences," *PLoS Comput. Biol.*, vol. 15, no. 6, 2019, Art. no. e1007129.

[17] Z.-H. Chen, Z.-H. You, Z.-H. Guo, H.-C. Yi, G.-X. Luo, and Y.-B. Wang, "Prediction of drug–target interactions from multi-molecular network based on deep walk embedding model," *Front. Bioeng. Biotechnol.*, vol. 8, 2020, Art. no. 338.

[18] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018. [Online]. Available: https://academic.oup.com/bioinformatics/article/34/17/i821/5093245

[19] M. Li, Z. Lu, Y. Wu, and Y. Li, "BACPIi: A bi-directional attention neural network for compound–protein interaction and binding affinity prediction," *Bioinformatics*, vol. 38, no. 7, pp. 1995–2002, 2022.

[20] X. Ru, Q. Zou, and C. Lin, "Optimization of drug–target affinity prediction methods through feature processing schemes," *Bioinformatics*, vol. 39, no. 11, 2023, Art. no. btad615.

[21] S. Zheng, Y. Li, S. Chen, J. Xu, and Y. Yang, "Predicting drug–protein interaction using quasi-visual question answering system," *Nature Mach. Intell.*, vol. 2, no. 2, pp. 134–140, 2020.

[22] L. Chen et al., "TransformerCPI: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," *Bioinformatics*, vol. 36, no. 16, pp. 4406–4414, 2020.

[23] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: Predicting drug–target binding affinity with graph neural networks," *Bioinf.*, vol. 37, no. 8, pp. 1140–1147, 2021.

[24] H. Wu et al., "AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism," *Neural Netw.*, vol. 169, pp. 623–636, 2024.

[25] X. Ru, X. Ye, T. Sakurai, and Q. Zou, "NerLTR-DTA: Drug–target binding affinity prediction based on neighbor relationship and learning to rank," *Bioinformatics*, vol. 38, no. 7, pp. 1964–1971, 2022.

[26] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drug–target interactions: A brief review," *Brief. Bioinf.*, vol. 15, no. 5, pp. 734–747, 2014.

[27] Y. Luo et al., "A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information," *Nature Commun.*, vol. 8, no. 1, pp. 1–13, 2017.

[28] Y. Ding, J. Tang, F. Guo, and Q. Zou, "Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization," *Brief. Bioinf.*, vol. 23, no. 2, 2022, Art. no. bbab582.

[29] M. Hao, S. H. Bryant, and Y. Wang, "Predicting drug-target interactions by dual-network integrated logistic matrix factorization," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017.

[30] H. Eslami Manoochehri and M. Nourani, "Drug-target interaction prediction using semi-bipartite graph model and deep learning," *BMC Bioinf.*, vol. 21, Jul. 2020, Art. no. 248. [Online]. Available: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3518-6

[31] Y. Shang, L. Gao, Q. Zou, and L. Yu, "Prediction of drug-target interactions based on multi-layer network representation learning," *Neurocomputing*, vol. 434, pp. 80–89, 2021.

[32] H. Fu, F. Huang, X. Liu, Y. Qiu, and W. Zhang, "MVGCN: Data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks," *Bioinformatics*, vol. 38, no. 2, pp. 426–434, 2022.

[33] Z. Chu et al., "Hierarchical graph representation learning for the prediction of drug-target binding affinity," *Inf. Sci.*, vol. 613, pp. 507–523, 2022.

[34] Y. Li, G. Qiao, K. Wang, and G. Wang, "Drug–target interaction predication via multi-channel graph neural networks," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab346.

[35] A. Dehghan, P. Razzaghi, K. Abbasi, and S. Gharaghani, "TripleMultiDTI: Multimodal representation learning in drug-target interaction prediction with triplet loss function," *Expert Syst. Appl.*, 2023, Art. no. 120754.

[36] Y. Wu, M. Gao, M. Zeng, J. Zhang, and M. Li, "BridgeDPI: A novel graph neural network for predicting drug–protein interactions," *Bioinformatics*, vol. 38, no. 9, pp. 2571–2578, 2022.

[37] A. Loukas, "How hard is to distinguish graphs with graph neural networks?," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 3465–3476.

[38] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proc. Int. Conf. Learn. Representations*, 2018.

[39] X. He and T.-S. Chua, "Neural factorization machines for sparse predictive analytics," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 355–364.

[40] K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun, "DeepPurpose: A deep learning library for drug–target interaction prediction," *Bioinformatics*, vol. 36, no. 22/23, pp. 5545–5547, 2020.

[41] Y. Su, R. Zhang, S. M. Erfani, and J. Gan, "Neural graph matching based collaborative filtering," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2021, pp. 849–858.

[42] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2014, pp. 1724–1734.

[43] M. I. Davis et al., "Comprehensive analysis of Kinase inhibitor selectivity," *Nature Biotechnol.*, vol. 29, no. 11, pp. 1046–1051, 2011.

[44] J. Tang et al., "Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis," *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 735–743, 2014.

[45] Q. Zhao, H. Zhao, K. Zheng, and J. Wang, "HyperAttentionDTI: Improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism," *Bioinformatics*, vol. 38, no. 3, pp. 655–662, 2022.

[46] G. Landrum et al., "RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling," *Greg Landrum*, vol. 8, no. 31.10, p. 5281, 2013.

[47] D. Osorio, P. Rondón-Villarreal, and R. Torres, "Peptides: A package for data mining of antimicrobial peptides," *Small*, vol. 12, pp. 44–444, 2015.

[48] K. Huang et al., "Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development," in *Proc. Neural Inf. Process. Syst., NeurIPS Datasets Benchmarks*, 2021.

[49] M. Tsubaki, K. Tomii, and J. Sese, "Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," *Bioinformatics*, vol. 35, no. 2, pp. 309–318, 2019.

[50] J. Wang, X. Li, and H. Zhang, "GNN-PT: Enhanced prediction of compound-protein interactions by integrating protein transformer," 2020, *arXiv:2009.00805*.

[51] W. Chen, G. Chen, L. Zhao, and C. Y.-C. Chen, "Predicting drug–target interactions with deep-embedding learning of graphs and sequences," *J. Phys. Chem. A*, vol. 125, no. 25, pp. 5633–5642, 2021.

[52] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2016.

[53] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, 1958, Art. no. 386.

[54] C. Zhang, Z. Yang, X. He, and L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 3, pp. 478–493, Mar. 2020.

[55] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. Annu. Conf. Neural Inf. Process. Syst., NIPS Workshop Deep Learn.*, 2014.

[56] X. Wang, H. Yang, and M. Zhang, "Neural common neighbor with completion for link prediction," in *Proc. 12th Int. Conf. Learn. Representations*, 2023.

[57] M. A. Subbaiah and N. A. Meanwell, "Bioisosteres of the phenyl ring: Recent strategic applications in lead optimization and drug design," *J. Med. Chem.*, vol. 64, no. 19, pp. 14046–14128, 2021.

[58] C. Lipinski, F. Lombardo, B. Dominy, and P. Feeney, "In vitro models for selection of development candidatesexperimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug. Del. Rev.*, vol. 23, no. 1, pp. 3–25, 1997.