DTI-ETA：基于GCN和GAT的异构图药物-靶标相互作用预测

DTI-HETA: prediction of drug-target interactions based on GCN and GAT on heterogeneous graph

Article *in* Briefings in Bioinformatics · April 2022

DOI: 10.1093/bib/bbac109

6 authors, including:

Song He
Institute of Health Service and Tranfusion Medicine
59 PUBLICATIONS **357** CITATIONS
SEE PROFILE

Xiao-Chen Bo
Beijing Institute of Radiation Medicine
352 PUBLICATIONS **3,676** CITATIONS
SEE PROFILE

OXFORD

Briefings in Bioinformatics, 2022, 1–16

https://doi.org/10.1093/bib/bbac109

Problem Solving Protocol

DTI-HETA:基于异构图GCN和GAT的药物-靶标相互作用预测

DTI-HETA: prediction of drug–target interactions based on GCN and GAT on heterogeneous graph

Kanghao Shao†, Yunhao Zhang†, Yuqi Wen†, Zhongnan Zhang, Song He and Xiaochen Bo

Corresponding authors: Zhongnan Zhang, Department of Software Engineering, School of Informatics, Xiamen University, Xiamen 361005, China. Fax: +86-592-2580500; E-mail: zhongnan_zhang@xmu.edu.cn; Song He, Beijing Institute of Radiation Medicine, Beijing 100850, China. E-mail: hes1224@163.com; Xiaochen Bo, Beijing Institute of Radiation Medicine, Beijing 100850, China. E-mail: boxiaoc@163.com

†These authors contributed equally to this work.

Abstract

Drug–target interaction (DTI) prediction plays an important role in drug repositioning, drug discovery and drug design. However, due to the large size of the chemical and genomic spaces and the complex interactions between drugs and targets, experimental identification of DTIs is costly and time-consuming. In recent years, the emerging graph neural network (GNN) has been applied to DTI prediction because DTIs can be represented effectively using graphs. However, some of these methods are only based on homogeneous graphs, and some consist of two decoupled steps that cannot be trained jointly. To further explore GNN-based DTI prediction by integrating heterogeneous graph information, this study regards DTI prediction as a link prediction problem and proposes an end-to-end model based on HETerogeneous graph with Attention mechanism (DTI-HETA). In this model, a heterogeneous graph is first constructed based on the drug–drug and target–target similarity matrices and the DTI matrix. Then, the graph convolutional neural network is utilized to obtain the embedded representation of the drugs and targets. To highlight the contribution of different neighborhood nodes to the central node in aggregating the graph convolution information, a graph attention mechanism is introduced into the node embedding process. Afterward, an inner product decoder is applied to predict DTIs. To evaluate the performance of DTI-HETA, experiments are conducted on two datasets. The experimental results show that our model is superior to the state-of-the-art methods. Also, the identification of novel DTIs indicates that DTI-HETA can serve as a powerful tool for integrating heterogeneous graph information to predict DTIs.

Keywords: DTI prediction, heterogeneous graph, graph neural network, graph attention network, link prediction

Introduction

Though many advances have been made in pharmaceutical research and development, the traditional drug discovery process is still risky, time-consuming and costly [1, 2], with a cost for a new molecular entity estimated at ∼12 years and $1.8 billion [1, 3]. Currently, the key in accelerating the drug discovery process is to determine whether a drug can interact with a target [4]. On one hand, identification of interactions between drugs and targets helps to effectively screen new drugs candidates [5]. Although there are tens of thousands of compounds stored in various databases, most compounds have no corresponding target information. With the growth of available data on drugs and targets, more and more scholars have tried to investigate effective computational methods to identify new drug–target interactions (DTIs). The traditional computational methods can be divided into three categories: docking simulation methods [6], ligand-based methods [7] and literature text mining [8]. Recently, some researchers have developed some predictive models for DTI based on machine learning, deep learning and network, which further expands the field and direction of DTI research [9–11]. Such methods consider not only the association between drugs but also the association between targets and often achieve desirable prediction results [12]. The discovery of new DTI facilitates the development of compounds into new medications. In addition, DTI prediction can aid drug repositioning. Identifying new indications or targets for existing drugs, namely drug repositioning, is another critical part in drug discovery [13]. With our understanding of pharmacology deepening, polypharmacology has been widely accepted [14]. Drugs often target multiple targets rather than a single target. Additionally, the same disease often involves multiple targets [15]. Such polypharmacological features accelerate the development of drug repositioning. As most of the approved drugs have been strictly verified for their safety, repositioned drugs can enter the clinical phase faster than new drugs [16]. Therefore, drug repositioning can significantly speed up the drug development process [5].

Kanghao Shao is a master student at the Xiamen University, Xiamen, China.
Yunhao Zhang is a PhD candidate at the Xiamen University, Xiamen, China.
Yuqi Wen is a PhD candidate at the Beijing Institute of Radiation Medicine, Beijing, China.
Zhongnan Zhang is a professor at the Xiamen University, Xiamen, China.
Song He is an associate professor at the Beijing Institute of Radiation Medicine, Beijing, China.
Xiaochen Bo is a professor at the Beijing Institute of Radiation Medicine, Beijing, China.

Received: December 27, 2021. **Revised:** February 14, 2022. **Accepted:** March 3, 2022

To sum up, DTI prediction is crucial for both the discovery of new drugs and the repositioning of existing drugs. Meanwhile, DTI prediction has become an important prerequisite in many cognate fields such as drug side-effect prediction, drug combination prediction and drug resistance study [14].

Considering the large chemical and genomic spaces and the complex interactions between drugs and targets, identifying DTIs by *in vivo* and *in vitro* experiments is still costly and time-consuming. To address this problem, recently developed computational prediction methods become indispensable technologies, and new methods are increasingly needed. Computational DTI prediction can benefit both the narrowing of the broad search space for candidate drugs for downstream laboratory experiments and the acceleration of new drug development [16, 17].

Currently, there are three main categories of computational methods for DTI prediction: ligand-based methods, docking simulations and chemogenomic methods [18]. While biologically well accepted, ligand-based methods and docking simulations are faced with many limitations, the number of known ligands is insufficient and the 3D structure of the proteins is unknown [5, 19]. The chemogenomic methods can be further classified into a few classes, such as machine learning-based methods and similarity-based methods [20]. Among these chemogenomic methods, the models based on machine learning and deep learning have received the most attention because of their reliable prediction results [5, 21–26]. In these methods, the knowledge about drugs and targets is encoded into features to train a model. Then the trained model is used to predict novel DTIs.

These methods usually involve both feature extraction and DTI prediction, but the potentially valid interaction of drug–target pairs is rarely considered in the construction of the model and cannot take advantage of the drug–drug and target–target similarity relationships [11, 27–29]. Also, this type of method only uses the DTI matrix as a binary label matrix for training, and the information contained in the heterogeneous biological data is ignored.

In recent years, the rapid development of the graph neural network (GNN) has extended the application of deep learning to the graph domain, and related methods have also been applied to drug discovery [30–32]. Network-based methods broadly consist of two steps: network construction and DTI prediction. Such methods consider not only the relationships among drugs but also the relationships among targets. However, the existing methods are designed for homogeneous graphs [33–36]. In reality, the drug data and target data have multiple data sources [37–40]. By integrating various information from heterogeneous data sources, the accuracy of DTI prediction can be further improved [38, 40, 41]. The heterogeneous network can encode diverse inter- and intra-relations between objects and has attracted increasing attention in recent years [42]. Sun *et al.* used symmetric meta-paths to obtain heterogeneous information and to calculate the similarity between nodes [43]. Dong *et al.* [44] obtained node sequences by meta-paths and used the graph embedding methods based on meta-paths to obtain node embeddings in heterogeneous graphs. Fan *et al.* [45] obtained node embeddings guided by meta-paths and used them for downstream recommendation tasks. Recently, researchers have attempted to use GNNs to analyze heterogeneous graphs. Schlichtkrull *et al.* [46] introduced graph convolutional neural (GCN) networks in the relationship modeling process to accomplish for node classification. Wang *et al.* [47] introduced an attention mechanism in the heterogeneous graph. Zhang *et al.* [48] proposed a model that can handle heterogeneous graphs with different attributes. Liao *et al.* [49] used latent features and attributes to learn node embeddings into graphs. Yun *et al.* [50] proposed Graph Transformer Networks (GTNs) to obtain node embedding representations in heterogeneous graphs and used them for downstream tasks.

However, these methods are easy to cause partial information loss during the data integration process, and they fail to consider the contribution of different neighboring nodes in aggregating central node information, which leads to poor prediction performance. Since separating feature learning from the prediction task may not yield the optimal solution, the prediction model should be trained through an end-to-end manner. An end-to-end model requires a large amount of data to understand the complex relationship between the input and the target. Meanwhile, the class imbalance between positive and negative data in the training set is also a challenge to GNN methods for DTI prediction [51].

DTI prediction generally includes three tasks: interactions between known drugs and targets, interactions between known drugs and new targets and interactions between new drugs and known targets. Our research aims to predict interactions between known drugs and targets, i.e. identifying new therapeutic opportunities for existing drugs. By considering DTI prediction as a link prediction problem on a heterogeneous graph, this study proposes a new prediction model called DTI-HETA, an end-to-end model based on HETerogeneous graph with Attention mechanism. In this model, a heterogeneous graph is first constructed based on the given drug–drug and target–target similarity matrices and the DTI matrix. Then, the GCN network is used to obtain the embedded representation of drugs and targets. Meanwhile, the Graph Attention Networks (GAT) based attention mechanism is adopted to highlight the contribution of different neighborhood nodes to the central node in aggregating the graph convolution information [52]. Finally, according to the embedded representation of the drugs and targets, a suitable decoder is selected for prediction. The main contributions of this study are as follows:

- A graph convolution strategy is designed for a heterogeneous graph to make full use of the information carried by the source dataset.

- The GAT is applied to highlight the contributions of neighboring nodes.
- The proposed model is trained in an end-to-end manner, and the model parameters can be updated more appropriately.

This study uses two datasets to evaluate the performance of the proposed model and compares it with some state-of-the-art models. The experimental results show that DTI-HETA achieves better prediction performance. Moreover, an in-depth literature survey is conducted to investigate the top predicted DTIs, and it is found that some of them have been supported by previous studies. Taken together, the results indicate that our model has an excellent ability for DTI prediction and provides a promising way to better understand the mode of drug action and drug repurposing.

Methodology

In this study, DTI prediction is regarded as a link prediction problem, that is, by judging whether there is an edge between the drug node and the target node and whether there is an interaction between the corresponding two instances can be predicted.

Definition 1: $G = (V, E, \mathcal{A}, \epsilon)$ is a graph, where V is the set of N nodes $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and E is the set of edges among different nodes. $\mathcal{A}$ and ϵ denote the set of node types and edge types, respectively. When $|\mathcal{A}| + |\epsilon| > 2$, G is a heterogeneous graph [48].

提出的模型由三部分组成：图结构、图嵌入和链路预测

The proposed model consists of three parts: graph construction, graph embedding and link prediction. First, a heterogeneous graph G is constructed based on the input drug–drug and target–target similarity matrices and the DTI matrix. In graph G, $|\mathcal{A}|$ equals to 2 and $|\epsilon|$ equals to 3, as shown in Figure 1. The node feature matrix is obtained by random initialization and graph embedding is used to obtain the embedding representation of drugs and targets based on GCN.

Definition 2: Node embedding in heterogeneous graphs. Given a heterogeneous graph G, the purpose of node embedding is to learn a function f that maps each node in G to a low-dimensional space $\mathbb{R}^d : f : v \in V \rightarrow \mathbb{R}^d$, where $d \ll |V|$.

映射的节点维度要远小于节点的个数

Considering that the neighboring nodes have different contributions to the central node in the aggregation process, this study introduces the GAT to obtain more meaningful node embeddings. Finally, the inner product decoder is used to predict the DTI based on the embedding representations obtained in the second step. The proposed model is trained in an end-to-end manner, and the parameters in the model are updated through gradient descent to minimize the loss function. This end-to-end training method is more likely to find effective models and embeddings for specific problems. The entire workflow of the proposed model is shown in Figure 2.

GCN on heterogeneous graphs

GCN is an important component of GNN [53]. Compared to the separation of embedding representation and the downstream tasks in graph embedding, GCN obtains the low-dimensional vector embedding of nodes, and then it performs end-to-end training according to different tasks, such as node classification, graph classification and link prediction [54].

In this study, the graph convolution module uses the neighboring nodes of the central node in graph G to define the information dissemination framework, which is called the node's local calculation graph here. The parameters and weights are shared among all local calculation graphs, and the same information propagating method should be used within the same local calculation graph. As shown in Figure 2, there are four different local calculation graphs: (a), (b), (c) and (d). In (a), the central node is a drug d_1, and all its neighboring nodes are drugs; in (b), the central node d_3 is a drug, and there are two types of neighboring nodes: drugs d_1 and d_5 and target t_4. (c) and (d) are another two cases where a target node is in the center. The features of the same drug node calculated by (a) and (b) are added to obtain its embedded representation. Similarly, according to (c) and (d), the feature representation of the target node can be obtained. The node embedding is calculated as follows:

hd表示节点d，had表示节点d在图(a)中的隐藏状态

$$h_d = h_d^{(a)} + h_d^{(b)}, h_t = h_t^{(c)} + h_t^{(d)}, \quad (1)$$

d表示药物，t表示靶标　　（1）：药物节点和靶标节点特征更新的公式

where h_d represents the embedding representation of drug node d; $h_d^{(a)}$and $h_d^{(b)}$represent the hidden states of node d in the local calculation graphs (a) and (b), respectively; h_t represents the embedding representation of the target node t and $h_t^{(c)}$ and $h_t^{(d)}$ represent the hidden states of the node t in the local calculation graphs (c) and (d), respectively.

In each layer of GCN, four local calculation graphs are calculated according to the type of edges in the original graph to propagate and aggregate node information. The aggregation method of single-layer graph convolution is shown in Equation (2):

在异构图中，中心节点i聚集其一阶邻居节点信息的聚集公式

$$h_i^{(k+1)} = \delta\left(\sum_{\tau}\sum_{j\in\mathcal{N}_\tau^i} W_\tau^{(k)} h_j^{(k)}\right), \quad (2)$$

where $h_i^{(k)} \in \mathbb{R}^{d^{(k)}}$ represents the hidden state of node i in the k-th layer of the GCN, and $d^{(k)}$represents the dimension of node embedding in the k-th layer. τ represents the

τ：表示边的类型
W(k)τ：表示边类型为τ的权重，且是在第k层卷积的权重。
Niτ：表示节点i在边类型为τ情况下的直接邻居节点的集合。

药物-药物之间的相似性具体指的是哪些方面的相似性？

包含药物副作用的相似性、药物化学结构的相似性、药物理化性质的相似性和药物治疗性质的相似性，四方面相似性

如何根据药物-药物之间的相似性、靶标-靶标之间的相似性、DTI矩阵构建异构图这异构图的“edge(边)”表示什么意思？

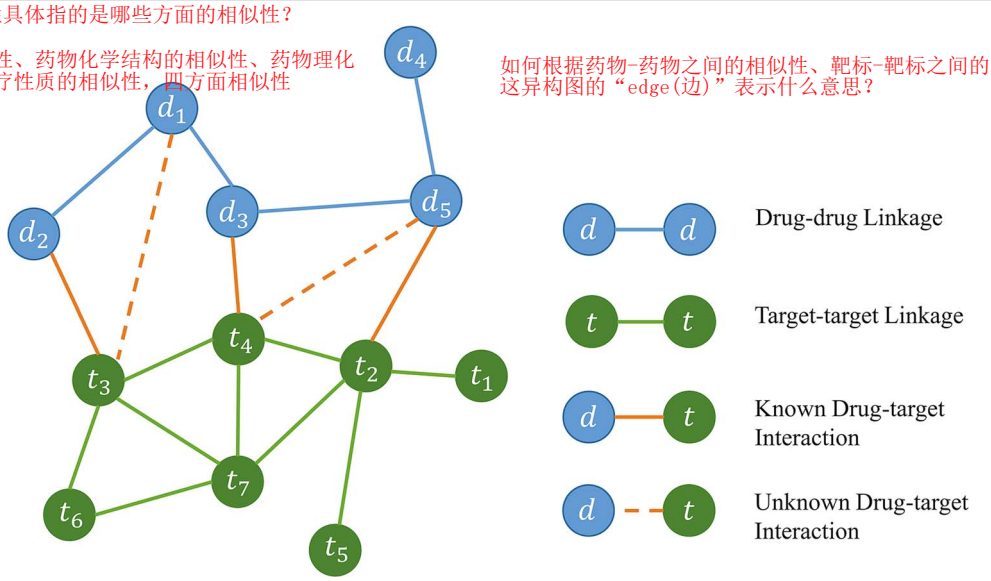

Figure 1. An example of heterogeneous graph.

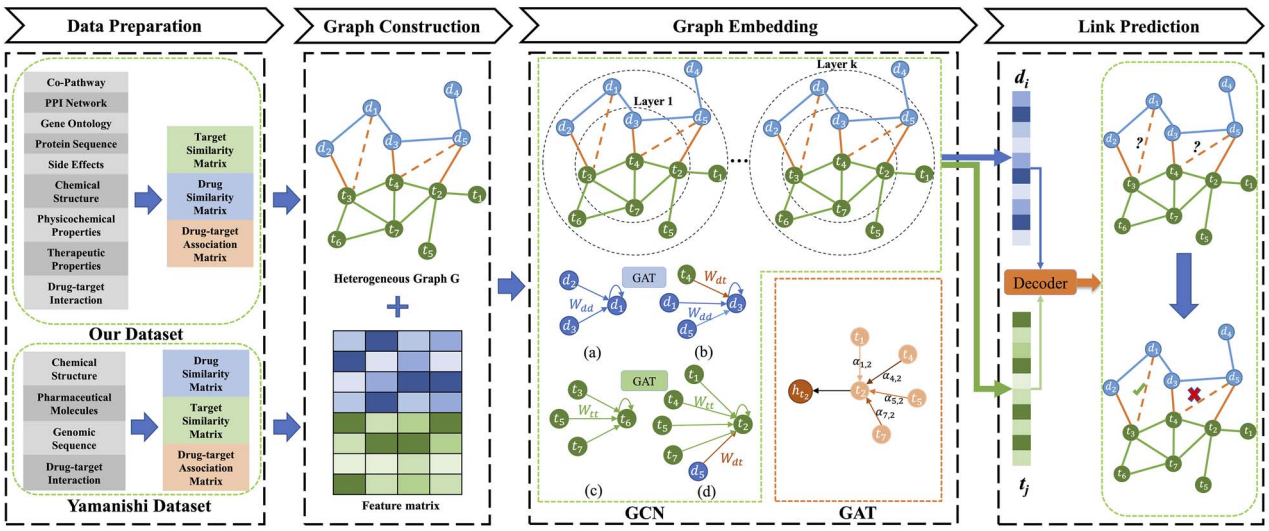

Figure 2. Detailed workflow of DTI-HETA including graph construction, graph embedding and link prediction.

Feature matrix是如何构建的？药物的特征具体是什么，靶标的特征又是什么？
药物的特征是指：一个药物分子和其他药物分子的相似性值构成的向量吗？

type of edge in the heterogeneous graph G, such as drug–drug (*dd*), target–target (tt) and drug–target (*dt*). $W_\tau^{(k)}$ is the weight of edge type τ in the k-th layer, and the weight of the same edge type is shared. $\mathcal{N}_\tau^i$ represents the set of direct neighbors of node i under type τ, including i itself. δ is the ReLU activation function.

Graph attention mechanism

GAT assigns different weights to neighboring nodes in the process of central node information aggregation. Taking nodes i and j as an example, GAT performs linear transformations on the two nodes respectively and then uses the mapping function f_a to assign attention coefficient e_{ij} to the nodes in the graph:

$$e_{ij} = f_a\left(W_\tau^{(k)} h_i^{(k)}, W_\tau^{(k)} h_j^{(k)}\right), \tag{3}$$

which represents the influence of node j on node i. In this study, f_a is a single-layer forward propagation neural network, and its parameter is a learnable vector $a_\tau^{(k)}$. The network is transformed with the LeakyRelu activation function. Therefore, e_{ij} can be calculated as follows:

$$e_{ij} = \sigma\left(a_\tau^{(k)}\left[W_\tau^{(k)} h_i^{(k)} \middle\| W_\tau^{(k)} h_j^{(k)}\right]\right), \tag{4}$$

where $\|$ represents the concatenation operator, and σ represents the LeakyRelu function.

To compare the attention coefficient between different nodes, the softmax function is used for normalization:

$$\alpha_{ij} = \text{softmax}\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{l \in \mathcal{N}_\tau^i} \exp\left(e_{il}\right)} \tag{5}$$

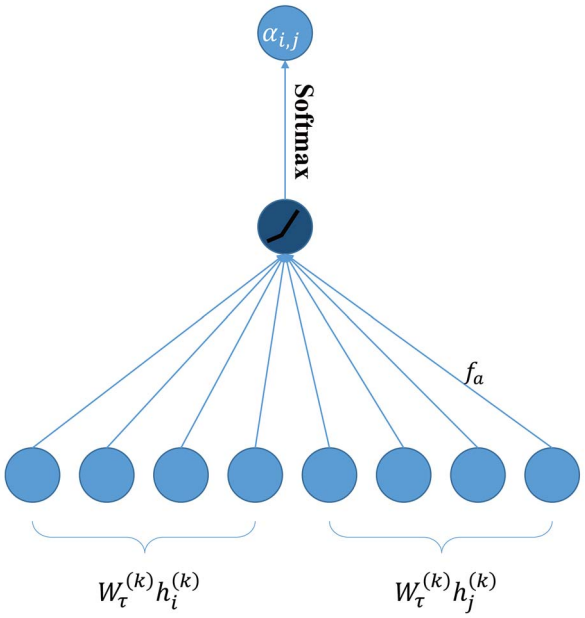

Figure 3. Schematic diagram of attention coefficient calculation [55].

The procedure for calculating the attention coefficient is shown in Figure 3.

After calculating the attention coefficient between nodes i and j, different weights (attentions) can be assigned to the neighboring nodes of the central node. The final output embedding of the central node i is the weighted summation of all nodes in $\mathcal{N}_\tau^i$. Thus, Equation (2) can be rewritten as

$$h_i^{(k+1)} = \delta\left(\sum_\tau \sum_{j \in \mathcal{N}_\tau^i} \alpha_{ij} W_\tau^{(k)} h_j^{(k)}\right). \quad (6)$$

Link prediction

In this study, an inner product decoder is used to predict the interaction between drug i and target j, and the following cross-entropy loss function is used to train the model and update the parameters in the model.

$$\begin{aligned} L\left(d_i, t_j\right) &= -Y\left(d_i, t_j\right) logP\left(d_i, t_j\right) \\ &\quad -\mathbf{E}_{t_n \sim \mathcal{D}\left(t_j\right)} \left(1 - Y\left(d_i, t_j\right)\right) \log\left(1 - P\left(d_i, t_n\right)\right) \end{aligned} \quad (7)$$

$Y(d_i, t_j) = 1$ indicates that there is an edge between the drug node i and the target node j. Based on known drug–target pairs (d_i, t_j) with interactions, a target t_n can be randomly selected from all targets that do not interact with drug d_i to form a negative sample (d_i, t_n). The negative samples used in the training process are generated by a negative sampling scheme. The final loss function is defined as follows:

$$L = \sum_{\left(d_i, t_j\right) \in \varepsilon_{dt}} L\left(d_i, t_j\right), \quad (8)$$

where ε_{dt} represents the set of drug–target edges in G.

Table 1. Size and data volume of two similarity matrices of our dataset after discretization

Matrix name	Size	No. of 1 s
Drug–drug similarity	580 × 580	16 820
Target–target similarity	2681 × 2681	359 388

Experiments

Data preparation

Two datasets are used in the experiments: our own dataset and the Yamanishi dataset [15]. In our own dataset, a drug–drug similarity matrix is obtained using Similarity Network Fusion (SNF) based on the similarities of drug side effects, drug chemical structure, drug physicochemical properties and therapeutic properties [56]. Meanwhile, a target–target similarity matrix is obtained using SNF based on the similarities of co-pathway, PPI network, Gene Ontology and gene-encoded protein sequence [57].

The drug–drug and target–target similarity matrices are dense matrices consisting of real numbers within [0, 1], and small elements indicate a weak similarity between drugs or targets. These two similarity matrices are discretized by setting appropriate thresholds to retain only the top 5% of the elements. These elements are set to 1, and the rest elements of the matrix are set to 0. The discretized matrices not only well retain the important similarity relationships in the original similarity matrix but also significantly reduce the complexity of the problem. The size and data volume of the two matrices after the discretization are shown in Table 1.

The drug–target associations in our dataset are extracted from DrugBank [58], a bioinformatics resource with high-quality drug data and accurate DTI data. As the gold standard, a 580 × 2681 drug–target association matrix with a total of 2187 known DTI s is obtained.

The Yamanishi dataset contains four subdatasets: Enzyme, Ion channel, G Protein-coupled Receptor (GPCR) and Nuclear receptor [15]. Each sub-dataset contains three networks: drug–drug structure similarity network, target–target similarity network and DTI network. Due to the small size of the GPCR and Nuclear receptor datasets, only the Enzyme and Ion channel sub-datasets are used in this study. The data volume of the Yamanishi dataset is shown in Table 2.

As a widely used dataset [59, 60], the Yamanishi dataset makes it easier for researchers to compare their algorithms with state-of-the-art methods. However, the

Table 2. Data volume of the Yamanishi dataset

	Enzyme	Ion channel
No. of drugs	212	99
No. of targets	478	146
No. of known interactions	1435	776

four protein families (Enzyme, Ion channel, GPCR and Nuclear receptor) contained in this dataset only account for 44% of the target molecules [61]. Also, the DTIs are divided into four types according to the protein families. This introduces data bias, which makes the prediction task easier (make predictions for only one type of DTIs). Meanwhile, the training model based only on these four types of DTIs cannot effectively predict other types of DTIs, thus reducing its generalization ability and practicality. Therefore, the Yamanishi dataset is usually used only in part of the performance evaluation.

Compared to the Yamanishi dataset, our dataset is larger and involves all the target types in DrugBank. In addition, our dataset fuses multiple types of drug and target features, covering various aspects of drug and target properties. The model trained on our dataset can be directly applied to predict novel DTIs.

For each dataset, 80% of the known DTI data are selected as the training set, 10% as the validation set and 10% as the test set. Ten-fold cross-validation is conducted to obtain the experimental results.

Model validation

The experiment includes three aspects. First, based on our own dataset, a set of optimal model parameters is obtained by analyzing the parameters involved in the model, such as the node embedding dimension and the number of graph convolution layers. Then, the effects of different decoders on the prediction performance are investigated. Finally, the model proposed in this study is compared with other methods on our own dataset and the Yamanishi dataset to verify its superiority.

In the process of model parameter optimization and model training, by considering the training time, the problem of pursuing the prediction effect while ignoring the time complexity can be avoided. In experiments of tuning the hyperparameters, the average time cost per epoch was set as the training time, so as to evaluate the time complexity for training the model and provide a reference for optimizing of the model parameters.

Accuracy, sensitivity, specificity, AUC, AUPR and Pre@K are six evaluation metrics commonly used in link prediction tasks. AUC indicates the area under the Receiver Operating Characteristic (ROC) curve, which is plotted with the true positive rate and false-positive rate in Equation (9) as the horizontal and vertical coordinates, respectively. The larger the value of AUC, the better the classification performance. AUPR indicates the area under the Precision-Recall (P–R) curve, which is plotted with recall and precision as the horizontal and vertical coordinates, respectively. The larger the value of AUPR, the better the model performance.

$$\text{TruePositiveRate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FalsePositiveRate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{9}$$

This study sorts the prediction results of the model and calculates the precision for the top-ranked K results, which is denoted as **Pre@K** and defined in (10):

$$Pre@K = \frac{|\, E_{\text{pred}}(1:K) \cap E_{\text{obs}} \,|}{K} \tag{10}$$

Adam optimizer is employed to train the model, and the learning rate is set to 0.001. To reduce overfitting, the dropout rate is set to 0.1 in the hidden layer, and regularization is added. Besides, the batch training strategy is adopted to reduce the memory overhead, and the batch size is set to 512.

The experimental environment is shown in Table 3.

Results and analysis

The effect of the node embedding dimension

In the model, the dimension of node embedding (d) and the number of layers of GCN (k) will affect the quality of node embedding representation. It can be seen from Figure 4 that with the increase of the node embedding dimension, the AUC and AUPR increase first and then decrease. This is because when d is too small, the obtained node embeddings do not carry enough information for the prediction task, which causes under-fitting and affects the final prediction performance; when d is too large, the extracted node embeddings may contain noise, which also affects the prediction performance.

In addition, the training time of the model increases almost exponentially with the dimension. Although there is some improvement in the prediction results when $d \geq 256$, considering the time cost of training, using a large embedding dimension is not beneficial. Therefore, it is more appropriate to set d to 32. In this case, better prediction results can be obtained with lower training costs.

The effect of the number of GCN layers

The effect of the number of graph convolution layers on the prediction performance is also investigated. It can be seen from Figure 5 that as the number of layers increases, both AUC and AUPR decrease. The purpose of graph convolution is to bring adjacent nodes closer, and increasing the number of layers infinitely will cause the node representation to converge to one point, resulting in excessive smoothness. Meanwhile, the increase in the number of layers also leads to a sharp increase in the number of model parameters, which may lead to overfitting and affect the prediction performance. Besides,

Table 3. Experimental environment

Hardware environment	**CPU Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz 10core GPU NVIDIA GeForce RTX 2080 RAM 128G**	
Software environment	Operating system	Ubuntu 16.04.4 LTS
	Program language	Python 3.7.1
	Libraries	Numpy 1.14.5 Scikit-learn 0.20.0 TensorFlow 1.8.0 Absl-py 0.2.2 Network 2.0 Tensorboard 1.8.0

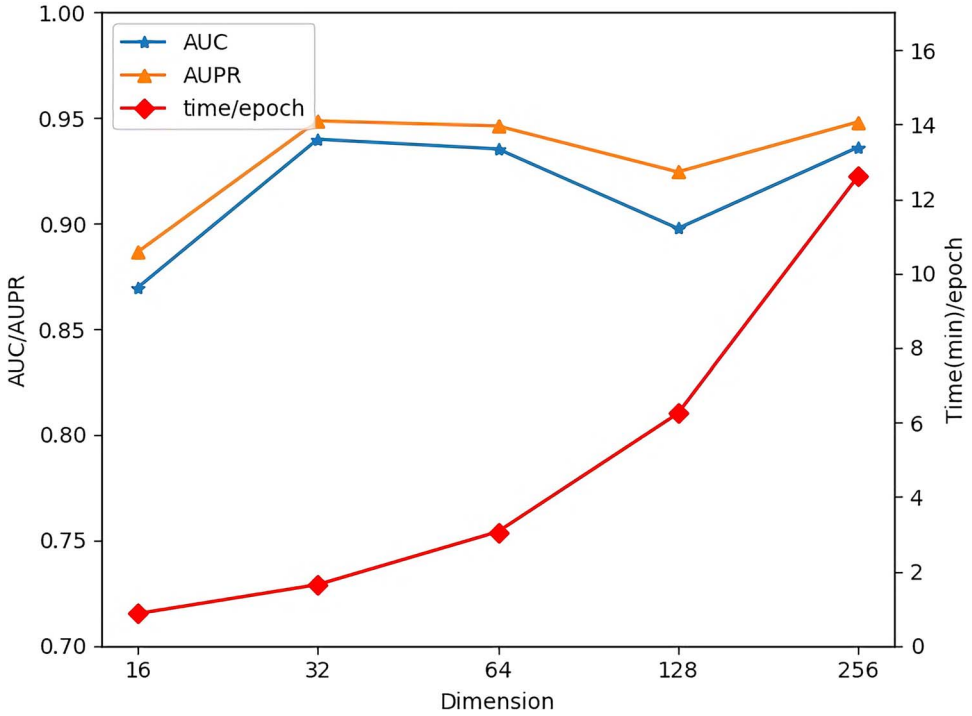

Figure 4. The influence of the dimension of node embedding on the prediction performance.

when the number of layers increases, the training time is gradually increasing. Therefore, when the number of graph convolutional layers is set to 1, better prediction performance can be obtained.

The effect of decoder

Figures 6 and 7 show the effects of using different decoders on the prediction performance of the model. In this experiment, the number of graph convolution layers is set to 1 and the node embedding dimension is set to 32. It can be seen from the two figures that the inner product decoder contributes to better prediction results than the bilinear decoder. This may be because the bilinear decoder introduces a trainable weight matrix, which increases the parameters in the model and causes overfitting.

In addition, the DTI-HETA model can obtain better prediction results even with a simple decoder, indicating that the model does not depend on a specific decoder.

Based on the above experimental results, the optimal parameter settings of the model can be obtained and used for subsequent experiments. These hyperparameters have been tuned and a set of optimal model parameters is obtained and summarized in Table 4.

Table 4. Summary of optimized parameter values

Parameters	**Value**
Learning rate	0.001
Dropout rate	0.1
Batch size	512
Dimension of node embedding	32
Number of layers of GCN	1

Performance comparison on our dataset

First, a comparison experiment is conducted on our dataset. Table 5 shows the comparison results between DTI-HETA and six other models on our dataset, and the best results are marked in boldface. These six models can be divided into two main categories: the models dealing with heterogeneous graph node embedding and relationship prediction separately, such as HIN2VEC [62],

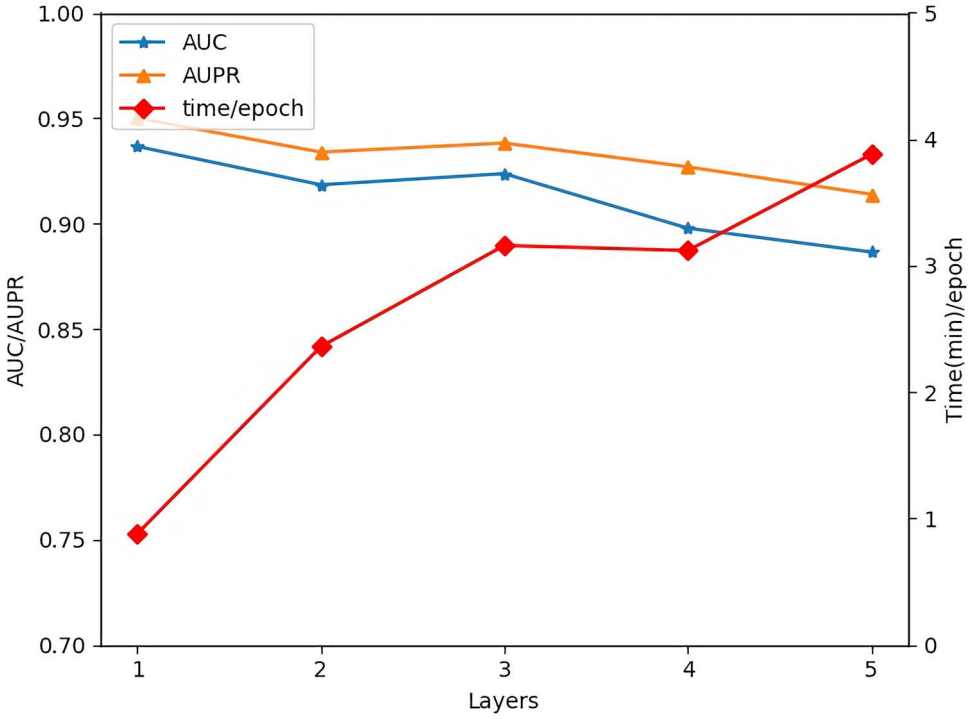

Figure 5. The influence of the number of graph convolutional layers on the prediction performance.

Table 5. AUC and AUPR comparison of different models on our dataset

	HEER	HIN2VEC	EVENT2VEC	JUST	GATNE	PGCN	DTI-HETA
AUC	0.7538	0.8553	0.8612	0.8530	0.8815	<u>0.8822</u>	**0.93224**
AUPR	0.7828	0.8870	0.8853	0.8725	0.8949	<u>0.9020</u>	**0.94722**

Best results are in boldface and the second best results are underlined.

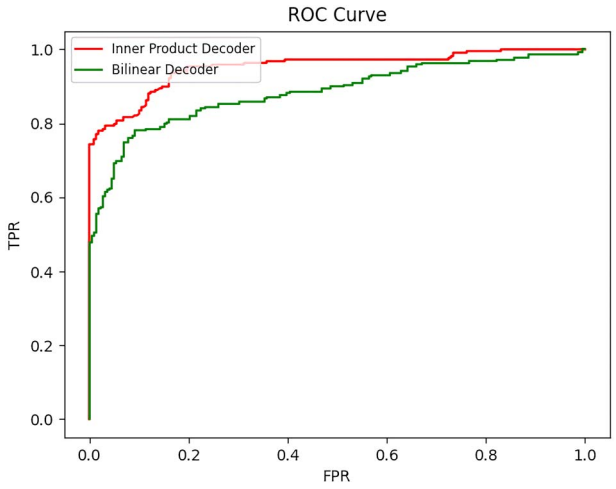

Figure 6. ROC curves with different decoders.

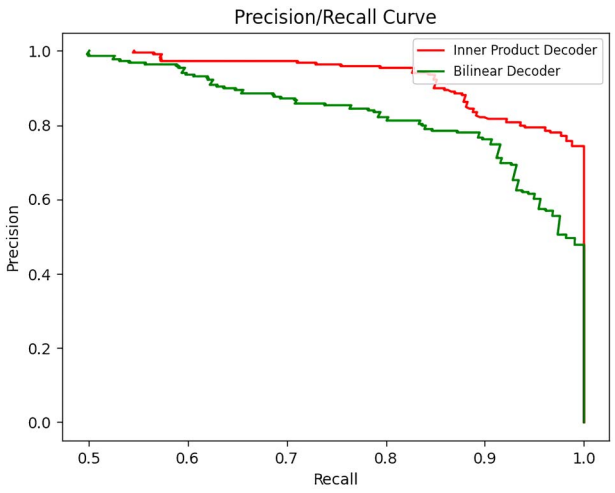

Figure 7. P–R curves with different decoders.

EVENT2VEC [63], JUST [64] and HEER [65], and those that are based on the end-to-end training and prediction, such as GATNE [66] and PGCN [67].

It can be seen from Table 5 that the DTI-HETA model achieves the best performance on our dataset according to AUC and AUPR. The AUC and AUPR of DTI-HETA are 5.004 and 4.522% higher than the corresponding second-best results, respectively. HIN2VEC, EVENT2VEC, JUST and HEER all consist of two parts: node embedding and link prediction. For these models, drug and target embeddings are first extracted from the heterogeneous graph, and then DTI s prediction is conducted based on the extracted embeddings. The end-to-end training of deep architectures generally provides better performance than training individual components separately. This is because the free parameters in all components

Table 6. AUC and AUPR comparison of different models on Yamanishi-Enzyme dataset

	GATNE	JUST	HIN2VEC	EVENT2VEC	IMCHGAN [70]	NormMulInf [70]	DTI-HETA
AUC	0.5220	0.6416	0.8613	0.9289	0.926	<u>0.958</u>	**0.96408**
AUPR	0.5278	0.5936	0.8960	<u>0.9449</u>	0.940	0.932	**0.97136**

Best results are in boldface and the second best results are underlined.

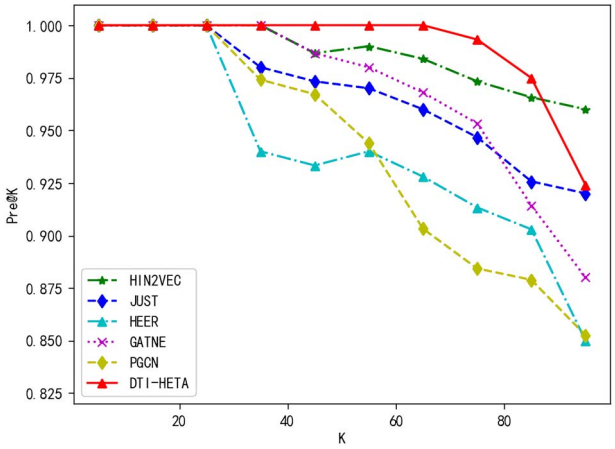

Figure 8. Pre@K comparison of different models on our dataset.

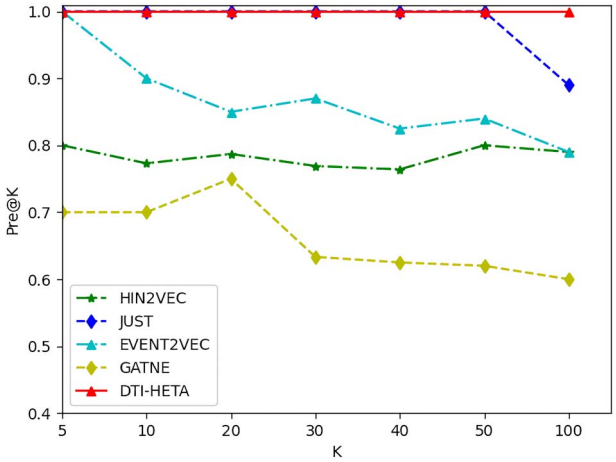

Figure 9. Pre@K comparison of different models on Yamanishi-Enzyme dataset.

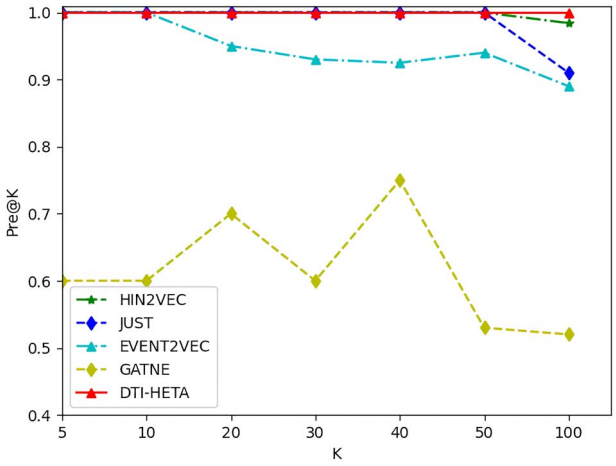

Figure 10. Pre@K comparison of different models on Yamanishi-Ion channel dataset.

can co-adapt and cooperate to achieve a single objective [68, 69]. Since these four models are not end-to-end models, their performances are not as good as that of GATNE and PGCN.

Figure 8 illustrates the Pre@K comparison of different models on our dataset. It can be seen that when $K < 30$, the Pre@K of all these methods is almost 1. With the increase of K, the performance of HEER becomes worse in terms of AUC and AUPR. The performance of PGCN and JUST also decreases significantly, while HIN2VEC achieves the second-best performance. Finally, when K is greater than 60, the performance of our model also decreases, but the number of false predictions has only increased by 1, 2 and 5 respectively when K is 75, 85 and 95.

Performance comparison on Yamanishi dataset

For further evaluation, the Yamanishi dataset is used for the performance evaluation of DIT-HETA. Since this dataset is widely used in other studies, some state-of-the-art models are selected for comparison, including IMCHGAN [70] and NormMulInf [71]. IMCHGAN is an end-to-end model and NormMulInf is not.

Tables 6 and 7 show the results of the comparison between DTI-HETA and other models in AUC and AUPR, and the best results are marked in boldface. It can be seen that DTI-HETA also achieves the best performance on both Yamanishi sub-datasets.

Taking the Yamanishi-Ion channel dataset as an example, the AUC and AUPR of DTI-HETA are 2.764 and 2.635% higher than the corresponding second-best results, respectively. Due to the end-to-end manner, IMCHGAN achieves relatively good results. NormMulInf utilizes collaborative filtering based on labeled and unlabeled interaction information. Compared with the two models, our model not only constructs a heterogeneous network to fully represent the information in the dataset but also exploits the attention mechanism for better feature representation learning. Therefore, our model achieves the best results.

The Pre@K of different models is also compared, and the results are illustrated in Figures 9 and 10. It can be seen that DTI-HETA significantly outperforms other models. Especially, DTI-HETA has a stable performance with the increase of K. GATNE has the worst performance, and other models show significant performance degradation when $K > 50$.

Table 7. AUC and AUPR comparison of different models on Yamanishi-Ion channel dataset

	GATNE	JUST	HIN2VEC	EVENT2VEC	IMCHGAN [70]	NormMulInf [70]	DTI-HETA
AUC	0.4952	0.6403	0.8530	0.9226	0.904	0.939	**0.96664**
AUPR	0.4298	0.5894	0.8992	0.9450	0.920	0.913	**0.97135**

Best results are in boldface and the second best results are underlined.

Table 8. Accuracy, sensitivity and specificity comparison of different models on Yamanishi-Enzyme dataset

	CnnDTI [27]	Zhao et al. [72]	PSSM+LPQ [73]	CNNEMS [74]	PreDTIs [75]	DTI-HETA
Accuracy	0.943	0.9032	0.8915	0.9419	0.9067	**0.94702**
Sensitivity	0.927	0.8903	0.8685	0.9191	0.9245	**0.94967**
Specificity	N/A	N/A	N/A	N/A	0.8578	**0.92219**

Best results are in boldface and the second best results are underlined.

Table 9. Accuracy, sensitivity and specificity comparison of different models on Yamanishi-Ion channel dataset

	CnnDTI [27]	Zhao et al. [72]	PSSM+LPQ [73]	CNNEMS [74]	PreDTIs [75]	DTI-HETA
Accuracy	0.919	0.8891	0.8601	0.9095	0.8989	**0.94**
Sensitivity	0.894	0.8961	0.8662	0.9031	0.9056	**0.936**
Specificity	N/A	N/A	N/A	N/A	0.8567	**0.944**

Best results are in boldface and the second best results are underlined.

Table 10. Two-tailed *P*-values of paired *t*-test on Yamanishi-Ion channel dataset

Paired model	AUC	AUPR
GATNE	2.08097E−15	3.17699E−17
JUST	5.68392E−14	3.71101E−16
HIN2VEC	9.79207E−08	3.25693E−08
EVENT2VEC	8.30856E−07	1.82298E−06
IMCHGAN	N/A	N/A
NormMulInf	N/A	N/A

The accuracy, sensitivity and specificity of different models are shown in Tables 8 and 9. Some state-of-the-art models are selected for comparison. Taking the Yamanishi-Ion Channel dataset as an example, the results predicted by different models are illustrated in Table 9. The accuracy, sensitivity and specificity of DTI-HETA are 2.1, 3.04 and 8.73% higher than the corresponding second-best results, respectively.

P-values are obtained by conducting a paired *t*-test between the AUC and AUPR values of the proposed DTI-HETA model and those of each comparable method. Taking Yamanishi-Ion channel dataset as an example, at $\alpha = 0.05$, the *P*-values of two-tailed test are presented for indicating statistical significance of improvements in Table 10. It should be noted that since only one set of data is available for the two referenced models (IMCHGAN and NormMulInf), no paired *t*-test experiments could be performed. The results indicate the statistical significance of improvements.

Visualization

To further illustrate the effectiveness of our model, visualization is performed on the testing dataset of Yamanishi-Enzyme. Based on the Hadamard product, a new embedding is generated for each drug–target pair. New embeddings are plotted in a two-dimensional space using t-SNE [76]. As shown in Figure 11, the green and gray dots respectively represent the drug–target pairs with (positive samples) and without interaction (negative samples). It can be seen that as the number of training epochs increases, the dots representing different types of samples are gradually distinguished, and finally the gray dots are basically separated. Due to the discretization in the link prediction, some gray dots are still closer to green dots, indicating that the decision boundary is more difficult to determine in link prediction.

The main reasons why DTI-HETA is superior to the above compared methods are summarized as follows:

- DTI-HETA constructs a heterogeneous network to fully represent the information in the input dataset;
- DTI-HETA is an end-to-end training model;
- DTI-HETA introduces a GAT in the process of node embedding, which can effectively obtain key features for the prediction task.

Identification of novel DTIs

Here, the novel DTIs predicted by DTI-HETA and not included in our training data set are analyzed. Specifically, the top 1% of all the prediction results are used in the following analysis.

To determine whether the predicted DTIs are in line with the current knowledge, the novel DTIs are compared with other existing DTI datasets collected from CTD [77], STITCH [78] and Matador [79].

The novel DTIs are first ranked according to the prediction results. Then, the number of overlapping pairs (between the predicted DTIs and the DTIs from other

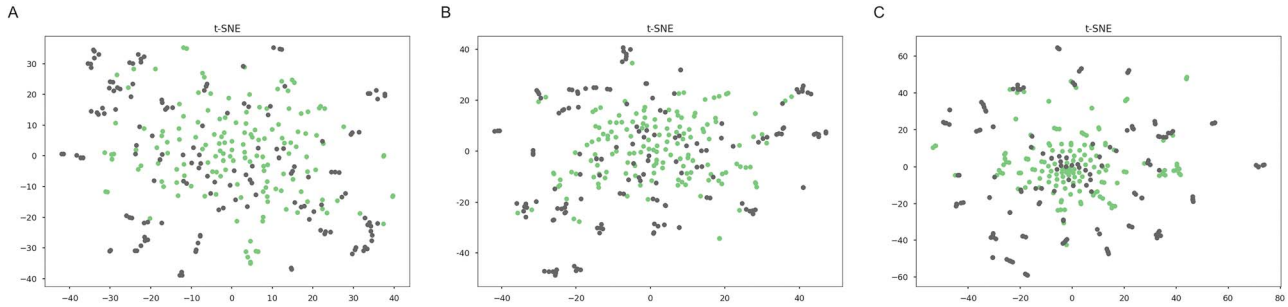

Figure 11. Visualization of the learned DTI embeddings on Yamanishi-Enzyme dataset. (A) Epoch 1; (B) Epoch 10; (C) Epoch 30.

Table 11. Enrichment of DTIs on other databases

	ES (known DTIs)	ES (predicted DTIs)	−log10 *P*-value (known DTIs)	−log10 *P*-value (predicted DTIs)
CTD	10.51	4.43	912.40	761.41
STITCH	26.98	17.90	98.14	156.72
Matador	46.55	9.32	481.29	115.50

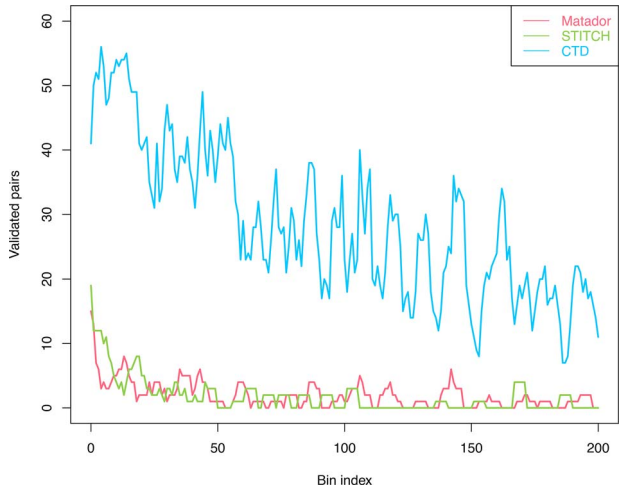

Figure 12. The overlap curves between predicted DTIs and known DTIs.

databases) is counted in the sliding bins of 500 consecutive interactions (Figure 12). The result indicates that the novel DTIs predicted by our model can be verified by other databases containing DTIs.

Furthermore, the enrichment score and *P*-value [80, 81] are used to quantify the appearance of predicted DTIs and known DTIs in other databases.

As shown in Table 11, the known DTIs and the predicted DTIs are both significantly enriched on other databases. The results indicate that the novel DTIs predicted by DTI-HETA have practical value and provide opportunities for drug repositioning.

Case study

Next, several cases are studied in the top50 novel DTIs predicted by DTI-HETA (Figure 13).

Apraclonidine, an alpha2-adrenergic agonist, is used for the treatment of raised intraocular pressure. The DrugBank database [82] indicates that apraclonidine can interact with adrenoceptor alpha 2A, adrenoceptor alpha 1A and adrenoceptor alpha 2B. Our new predictions indicate that apraclonidine can interact with adrenoceptor alpha 2C, an alpha-2 adrenergic receptor that plays a critical role in the central nervous system. A previous study, which demonstrated that apraclonidine can have a direct interaction with ADRA2C, can support this new prediction [83].

Dasatinib is a tyrosine kinase inhibitor used to treat chronic myeloid leukemia and acute lymphoblastic leukemia. The main targets of Dasatinib are BCR activator of RhoGEF and GTPase-Abl tyrosine kinase (BCR-ABL), SRC family, c-KIT, Erythropoietin-producing Hepatocellular (EPH) receptor A2 and platelet-derived growth factor receptor beta. Erb-b2 receptor tyrosine kinase 4 (ERBB4) is a member of the EGFR subfamily of receptor tyrosine kinases. Mutations in this gene have been associated with various cancer types including melanoma, lung adenocarcinoma and medulloblastoma [84]. Our new predictions indicate that ERBB4 may be a novel target of Dasatinib. This is consistent with the conclusion of a previous study that Dasatinib has a moderate affinity for ERBB4 [85].

Ibrutinib, an antineoplastic drug for treating mantle cell lymphoma, chronic lymphocytic leukemia, Waldenström's macroglobulinemia and chronic graft versus host disease, acts as an irreversible potent inhibitor of Burton's tyrosine kinase (BTK). Our new predictions demonstrate that Ibrutinib can act on Janus kinase 2 (JAK2), a tyrosine kinase that plays an important role in cytokine and growth factor signaling. During the initial characterization of ibrutinib, it was found that ibrutinib could bind to other kinases as well [86]. A previous study has shown that Ibrutinib enhances macrophage-mediated antibody-dependent cellular phagocytosis independent of BTK-inhibition via targeting of JAK2 [87]. This off-target effect of

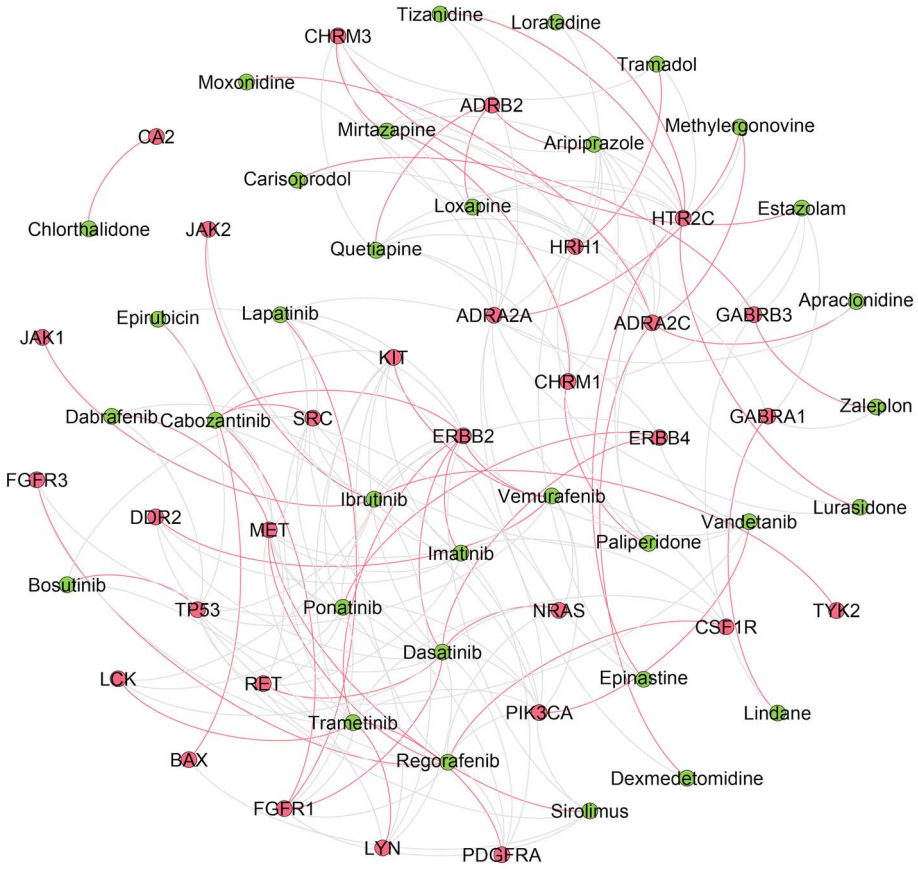

Figure 13. Network visualization of the top50 DTIs predicted by DTI-HETA. Targets are shown in red circles and drugs are shown in green circles. Known DTIs are shown by grey edges and the novel predicted DTIs are shown by red edges.

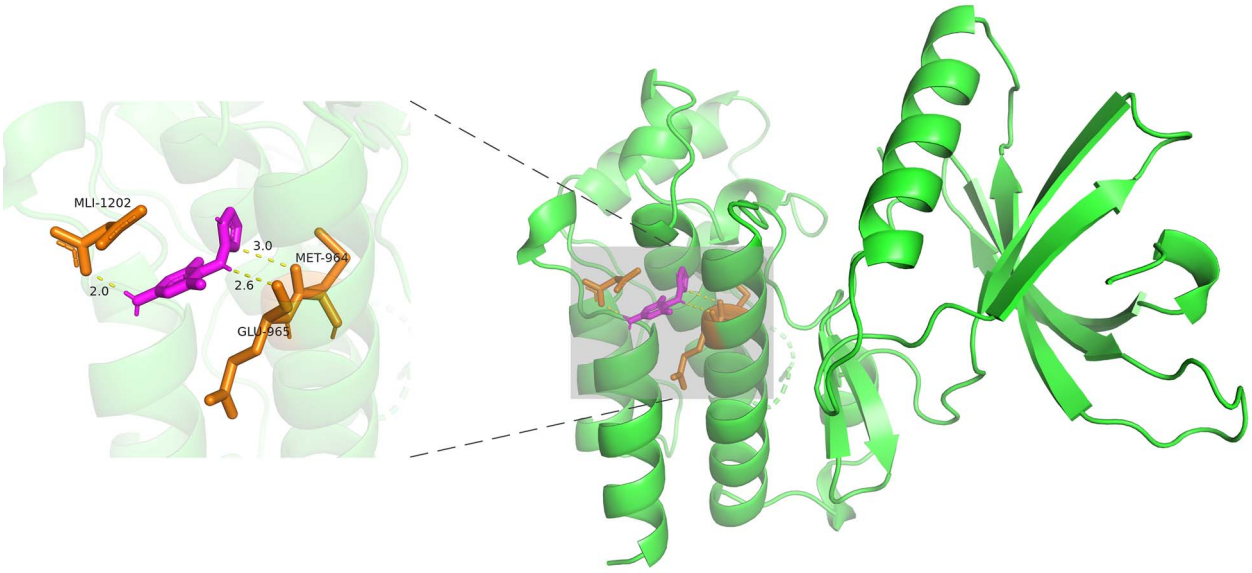

Figure 14. The docked pose between Ibrutinib and JAK2. The figure is pictured by PyMOL [89].

ibrutinib could be beneficial and may provide new indications for clinical applications. To further validate this new interaction, computational docking is conducted and the docking program AutoDock [88] is utilized to infer the possible binding pattern of the new predicted DTI. The docking result indicates that ibrutinib can dock to the structure of JAK2. More specifically, Ibrutinib binds to JAK2 (Figure 14) by forming hydrogen bonds with residues MLI1202, GLU965 and MET964.

The above cases illustrate that our model has an excellent ability for DTI prediction and provides hints for understanding the mode of drug action and drug repurposing.

Conclusion

Nowadays, the heterogeneous graph about drugs and target proteins has become a powerful tool in DTI prediction. The great potential of the GNN model based on a drug–target heterogeneous graph has not been fully exploited. The category imbalance of positive and negative data may affect the performance of the GNN approach. In addition, GNN may easily capture possible biased patterns in the dataset. In this study, we proposed DTI-HETA, an end-to-end GCN model for predicting the DTI from heterogeneous data sources. DTI-HETA obtains the node embedding representation of drugs and targets by defining the graph convolution that introduces the attention mechanism into the heterogeneous graph and then employs the decoder to predict potential DTIs. The experimental results indicate that DTI-HETA outperforms both state-of-the-art end-to-end models and non-end-to-end models. Meanwhile, the validation on external datasets and case studies shows that DTI-HETA can identify effective DTIs.

Although DTI-HETA demonstrates great prediction performance, this study is faced with some challenges. First, unknown drug–target pairs are randomly selected as negative samples, which may limit the prediction accuracy of the model. Experimentally measured negative samples will be continuously needed in the future. Another challenge is that more drug- and target-related heterogeneous network could be incorporated and further explored, such as metabolic network and drug–disease network. These various heterogeneous networks will provide rich semantic information that is conducive to DTI prediction.

In summary, DTI-HETA, a GCN-based model that exploits heterogeneous graph information, proves its capability to server as an applicable model for the identification of effective DTIs. Although it is used to predict DTIs in this study, DTI-HETA is scalable for heterogeneous graph integration and can be used to predict other types of associations, such as drug–drug interactions and protein–protein interactions. The accurate computational identification of DTIs can enhance the understanding of related biological processes and complicated biological interactions on the one hand and accelerate the process of drugs entering clinical trials by predicting new targets for the existing drugs on the other hand. Especially, the model is beneficial for researchers to experimentally validate predicted DTIs by narrowing the search space. At last, personalized medicine is an evolving field of medicine that provides personalized treatment for patients, so using a single or several drug targets for all the patients is inappropriate [90]. Combining disease marker-based networks or disease-specific networks with our model will be helpful for the discovery of personalized drugs. The prediction of personalized drug targets will have an important impact on personalized therapy and benefit the success of clinical trials [91]. Overall, by narrowing the search space of DTIs, DTI-HETA is a powerful model for the discovery of novel DTIs and may provide important hints for understanding the underlying mechanisms of drug action.

Key Points

- Design a predictive model for a heterogeneous graph to make full use of the information carried by the data.
- Design the corresponding graph convolution strategy for a heterogeneous graph and introduce GAT to highlight the different contributions of neighboring nodes.
- Train the model in an end-to-end manner, and the model parameters can be updated more pertinently.

Data Availability

https://github.com/ZhangyuXM/DTI-HETA

Funding

National Natural Science Foundation of China (No. 62103436 to S.H.).

References

1. Paul SM, Mytelka DS, Dunwiddie CT, *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;**9**(3):203–14.
2. Adams CP, Brantner VV. Estimating the cost of new drug development: is it really $802 million? *Health Aff* 2006;**25**(2):420–8.
3. Lotfi Shahreza M, Ghadiri N, Mousavi SR, *et al.* A review of network-based approaches to drug repositioning. *Brief Bioinform* 2018;**19**(5):878–92.
4. Núñez S, Venhorst J, Kruse CG. Target–drug interactions: first principles and their application to drug discovery. *Drug Discov Today* 2012;**17**(1–2):10–22.
5. Chen R, Liu X, Jin S, *et al.* Machine learning for drug-target interaction prediction. *Molecules* 2018;**23**(9):2208.
6. Cheng AC, Coleman RG, Smyth KT, *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* 2007;**25**(1):71–5.
7. Keiser MJ, Roth BL, Armbruster BN, *et al.* Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**(2):197–206.
8. Zhu S, Okuno Y, Tsujimoto G, *et al.* A probabilistic model for mining implicit 'chemical compound–gene'relations from literature. *Bioinformatics* 2005;**21**(suppl_2):ii245–51.
9. Shang ZW, Jin LI, Jiang YS, *et al.* A method of drug target prediction based on SVM and its application. *Progr Modern Biomed* 2012;**12**(20):3943–3946.
10. Yu H, Chen J, Xu X, *et al.* A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PLoS One* 2012;**7**(5):e37608.

11. Hu PW, Chan KCC, You ZH. Large-scale prediction of drug-target interactions from deep representations.. In: *International Joint Conference on Neural Networks (IJCNN)*. Vancouver, BC, Canada IEEE, 2016. pp.1236–43.
12. Buza K, Peška L. Drug–target interaction prediction with bipartite local models and hubness-aware regression. *Neurocomputing* 2017;**260**:284–93.
13. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, *et al.* Drug repositioning and repurposing: terminology and definitions in literature. *Drug Discov Today* 2015;**20**(8):1027–34.
14. Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH. Drug-target and disease networks: polypharmacology in the post-genomic era. *In Silico Pharmaco* 2013;**1**(1):1–4.
15. Yamanishi Y, Araki M, Gutteridge A, *et al.* Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**(13):i232–40.
16. Ru X, Ye X, Sakurai T, *et al.* Current status and future prospects of drug–target interaction prediction. *Brief Funct Genom* 2021;**20**(5): 312–22.
17. Chen X, Yan CC, Zhang X, *et al.* Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**(4):696–712.
18. Ezzat A, Wu M, Li XL, *et al.* Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform* 2019;**20**(4):1337–57.
19. Yıldırım MA, Goh KI, Cusick ME, *et al.* Drug—target network. *Nat Biotechnol* 2007;**25**(10):1119–26.
20. Sachdev K, Gupta MK. A comprehensive review of feature based methods for drug target interaction prediction. *J Biomed Inform* 2019;**93**:103159.
21. Chen X, Liu MX, Yan GY. Drug–target interaction prediction by random walk on the heterogeneous network. *Mol BioSyst* 2012;**8**(7):1970–8.
22. Pliakos K, Vens C. Drug-target interaction prediction with tree-ensemble learning and output space reconstruction. *BMC Bioinform* 2020;**21**(1):1–11.
23. Zeng X, Zhu S, Hou Y, *et al.* Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics* 2020;**36**(9):2805–12.
24. Bagherian M, Kim RB, Jiang C, *et al.* Coupled matrix–matrix and coupled tensor–matrix completion methods for predicting drug–target interactions. *Brief Bioinform* 2021;**22**(2):2161–71.
25. Wang J, Wang H, Wang X, *et al.* Predicting drug-target interactions via FM-DNN learning. *Curr Bioinforma* 2020;**15**(1): 68–76.
26. Cai J, Cai H, Chen J, *et al.* Identifying "many-to-many" relationships between gene-expression data and drug-response data via sparse binary matching. *IEEE/ACM Trans Comput Biol Bioinform* 2018;**17**(1):165–76.
27. Hu SS, Zhang C, Chen P, *et al.* Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks. *BMC Bioinform* 2019;**20**(25): 1–12.
28. Xie LW, He S, Song XY, *et al.* Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genomics* 2018;**19**(7):667.
29. Zheng X P, He S, Song X Y, *et al.* DTI-RCNN: new efficient hybrid neural network model to predict drug–target interactions. In: *International Conference on Artificial Neural Networks*. Springer, Cham, 2018. pp. 104–14.
30. Sun M, *et al.* Graph convolutional networks for computational drug development and discovery. *Brief Bioinform* 2020;**21**(3): 919–35.
31. Abbasi K, Razzaghi P, Poso A, *et al.* DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics* 2020;**36**(17):4633–42.
32. Zhang Z, Chen L, Zhong F, *et al.* Graph neural network approaches for drug-target interactions. *Curr Opin Struct Biol* 2022;**73**:102327.
33. Zhou C, Liu Y, Liu X, *et al.* Scalable graph embedding for asymmetric proximity. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, San Francisco, California USA. AAAI, 2017, pp. 2942–2948.
34. Zhao X, Chang A, Sarma AD, *et al.* On the embeddability of random walk distances. *Proc VLDB Endow* 2013;**6**(14): 1690–701.
35. Wang X, Cui P, Wang J, *et al.* Community preserving network embedding. In: *Thirty-first AAAI Conference on Artificial Intelligence*, San Francisco, California, USA. 2017, pp 203–209.
36. Ribeiro L F R, Saverese P H P, Figueiredo D R. struc2vec: Learning node representations from structural identity. In: *Proceedings of the 23rd ACM SIGKDD International Conference .on Knowledge Discovery and Data Mining*, Halifax, NS, Canada. ACM, New York, NY, United States. 2017. pp. 385–94.
37. Zheng X, Ding H, Mamitsuka H, *et al.* Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, Illinois, USA. ACM, New York, NY, United States. 2013. pp. 1025–33.
38. Luo Y, Zhao X, Zhou J, *et al.* A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):1–13.
39. Nagarajan N, Dhillon IS. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics* 2014;**12**:i60–8.
40. Wan F, Hong L, Xiao A, *et al.* NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* 2019;**35**(1):104–11.
41. Zheng, Xiaodong, *et al.* "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013.
42. Yu G, Wang Y, Wang J, *et al.* Attributed heterogeneous network fusion via collaborative matrix tri-factorization. *Inf Fusion* 2020;**63**:153–65. https://doi.org/10.1016/j.inffus.2020.06.012.
43. Sun Y, Han J, Yan X, *et al.* Pathsim: meta path-based top-k similarity search in heterogeneous information networks. *Proc VLDB Endow* 2011;**4**(11):992–1003.
44. Dong Y, Chawla N V, Swami A. metapath2vec: scalable representation learning for heterogeneous networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on knowledge Discovery and Data Mining*, Halifax, NS, Canada. ACM, New York, NY, United States. 2017. pp. 135–44.
45. Fan S, Zhu J, Han X, *et al.* Metapath-guided heterogeneous graph neural network for intent recommendation. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA. ACM, New York, NY, United States. 2019. pp. 2478–86.
46. Schlichtkrull M, Kipf T N, Bloem P, *et al.* Modeling relational data with graph convolutional networks. In: *European Semantic Web Conference*, Springer, Cham, 2018. pp. 593–607.
47. Wang X, Ji H, Shi C, *et al.* Heterogeneous graph attention network. In: *The World Wide Web Conference*, San Francisco, CA, USA. ACM, New York, NY, United States. 2019. pp. 2022–32.
48. Zhang C, Song D, Huang C, *et al.* Heterogeneous graph neural network. In: *Proceedings of the 25th ACM SIGKDD. International*

Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA. ACM, New York, NY, United States. 2019. pp. 793–803.
49. Liao L, He X, Zhang H, *et al.* Attributed social network embedding. *IEEE Trans Knowl Data Eng* 2018;**30**(12):2257–70.
50. Yun S, Jeong M, Kim R, *et al.* Graph transformer networks. In: *Advances in Neural Information Processing Systems*, 2019, 11983–93.
51. Lim J, Ryu S, Park K, *et al.* Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J Chem Inf Model* 2019;**59**(9):3981–8.
52. Veličković P, *et al.* Graph attention networks. *arXiv preprint* 2017;arXiv:1710.10903.
53. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks[J]. *Bioinformatics*, 2018;**34**(13):i457–i466.
54. Shanthamallu U S, Thiagarajan J J, Spanias A. Uncertainty-matching graph neural networks to defend against poisoning attacks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, held virtually. AAAI Press, Palo Alto, California, USA, 2021, Vol. **35**(11). pp. 9524–32.
55. Veličković P, Cucurull G, Casanova A, *et al.* Graph Attention Networks. In: *6th International Conference on Learning Representations*, 2018.
56. He S, Wen Y, Yang X, *et al.* PIMD: an integrative approach for drug repositioning using multiple characterization fusion. *Genom Proteom Bioinformat* 2020;**18**(5):565–81.
57. Wu LL, Wen YQ, Yang XX *et al.* Synthetic lethal interactions prediction based on multiple similarity measures fusion. *J Comput Sci Technol* **36**(2): 261–75 Mar. 2021.
58. Law V, Knox C, Djoumbou Y, *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**(D1): D1091–7.
59. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;**34**(17): i821–9.
60. Mahmud SMH, Chen W, Meng H, *et al.* Prediction of drug-target interaction based on protein features using undersampling and feature selection techniques with boosting. *Anal Biochem* Elsevier Inc 2020;**589**:113507.
61. Xu L, Ru X, Song R. Application of machine learning for drug–target interaction prediction. *Front Genet* 2021;**12**: 1077.
62. Fu, Tao-yang, Wang-Chien Lee, Zhen Lei. HIN2VEC: explore meta-paths in heterogeneous information networks for representation learning. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore. ACM, New York, NY, United States. 2017.
63. Chu Y, Feng C, Guo C, *et al.* Event2vec: heterogeneous hypergraph embedding for event data. In: *2018 IEEE International Conference on Data Mining Workshops (ICDMW)* Singapore *IEEE*, 2018. pp. 1022–9.
64. Hussein, Rana, Dingqi Yang, and Philippe Cudré-Mauroux. Are meta-paths necessary? Revisiting heterogeneous graph embeddings. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino, Italy. ACM, New York, NY, United States. 2018.
65. Shi, Yu, *et al.* Easing embedding learning by comprehensive transcription of heterogeneous information networks. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London, United Kingdom. ACM, New York, NY, United States. 2018.
66. Cen, Yukuo, *et al.* Representation learning for attributed multiplex heterogeneous network. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA. ACM, New York, NY, United States. 2019, pp. 1358–1368.
67. Li Y, *et al.* PGCN: disease gene prioritization by disease and gene embedding through graph convolutional neural networks. *bioRxiv* 2019;532226. doi: https://doi.org/10.1101/532226.
68. Mnih V, Kavukcuoglu K, Silver D, *et al.* Human-level control through deep reinforcement learning. *Nature* 2015;**518**(7540): 529–33.
69. Valmadre J, Bertinetto L, Henriques J, *et al.* End-to-end representation learning for correlation filter based tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, United States. IEEE Computer Society 2017. pp. 2805–13.
70. Li J, Wang J, Lv H, *et al.* IMCHGAN: inductive matrix completion with heterogeneous graph attention networks for drug-target interactions prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021. doi: 10.1109/TCBB.2021.3088614.
71. Peng L, Liao B, Zhu W, *et al.* Predicting drug–target interactions with multi-information fusion. *IEEE J Biomed Health Inform* 2015;**21**(2):561–72.
72. Zhao ZY, Huang WZ, Pan J, *et al.* A sparse feature extraction method with elastic net for drug-target interaction identification. *Sci Program* 2021;**2021**:6686409.
73. Li Y, Huang YA, You ZH, *et al.* Drug-target interaction prediction based on drug fingerprint information and protein sequence. *Molecules* 2019;**24**(16):2999.
74. Yan X, You Z H, Wang L, *et al.* CNNEMS: using convolutional neural networks to predict drug-target interactions by combining protein evolution and molecular structures information. In: *International Conference on Intelligent Computing*. Springer, Cham, 2021. pp. 570–9.
75. Mahmud SMH, Chen W, Liu Y, *et al.* PreDTIs: prediction of drug–target interactions based on multiple feature information using gradient boosting framework with data balancing and feature selection techniques. *Brief Bioinform* 2021;**22**(5): bbab046.
76. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**(11):2579–2605.
77. Davis AP, Grondin CJ, Johnson RJ, *et al.* The comparative toxicogenomics database: update 2019. *Nucleic Acids Res* 2019;**47**(D1):D948–54.
78. Szklarczyk D, Santos A, Von Mering C, *et al.* STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 2016;**44**(D1):D380–4.
79. Günther S, Kuhn M, Dunkel M, *et al.* SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res* 2007;**36**(suppl_1):D919–22.
80. Iorio F, Bosotti R, Scacheri E, *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci* 2010;**107**(33):14621–6.
81. Ye Y, Wen Y, Zhang Z, *et al.* Drug-target interaction prediction based on adversarial Bayesian personalized ranking. *Biomed Res Int* 2021;**2021**:6690154.
82. Wishart DS, Knox C, Guo AC, *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;**34**(suppl_1):D668–72.
83. Munk SA, Harcourt D, Ambrus G, *et al.* Synthesis and evaluation of 2-[(5-Methylbenz-1-ox-4-azin-6-yl) imino] imidazoline, a potent, peripherally acting $\alpha 2$ adrenoceptor agonist. *J Med Chem* 1996;**39**(18):3533–8.
84. Arteaga CL, Engelman JA. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* 2014;**25**(3):282–303.

85. Carter TA, Wodicka LM, Shah NP, *et al.* Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proc Natl Acad Sci* 2005;**102**(31):11011–6.
86. Berglöf A, Hamasy A, Meinke S, *et al.* Targets for ibrutinib beyond B cell malignancies. *Scand J Immunol* 2015;**82**(3): 208–17.
87. Barbarino V, Henschke S, Blakemore SJ, *et al.* Macrophage-mediated antibody dependent effector function in aggressive B-cell lymphoma treatment is enhanced by ibrutinib via inhibition of JAK2. *Cancer* 2020;**12**(8):2303.
88. Morris GM, Huey R, Lindstrom W, *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;**30**(16): 2785–91.
89. DeLano WL. The PyMOL molecular graphics system. http://wwwPymolOrg. 2002.
90. Wang E, Zou J, Zaman N, *et al.* Cancer systems biology in the genome sequencing era: part 2, evolutionary dynamics of tumor clonal networks and drug resistance. *Seminars in Cancer Biology*. Elsevier Ltd, 2013. Vol. **23**(4), pp. 286–92.
91. Wang E, Zaman N, Mcgee S, *et al.* Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Seminars in Cancer Biology*. Elsevier Ltd, 2015. Vol. **30**. pp. 4–12.