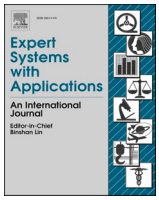
Expert Systems With Applications 249 (2024) 123730

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Drug-target interactions prediction via graph isomorphic network and cyclic training method

Yuhong Du [a,*], Yabing Yao [a], Jianxin Tang [a], Zhili Zhao [b], Zhuoyue Gou [c]

[a] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
[b] School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China
[c] Northwest Institute of Nuclear Technology, Xian 710024, China

ARTICLE INFO

Keywords:
Drug-target interactions
Graph isomorphic network
Cyclic training method

ABSTRACT

Predicting drug-target interactions through computational methods holds the potential to provide more reliable candidates for subsequent experimental validation and reduce associated costs. Most methods for Drug-target Interactions (DTIs) prediction have made advancements from two perspectives, improving the accuracy of drug and target representations, and seeking more precise mapping functions between the drug and target spaces. In this study, we propose a model called CT-GINDTI, which prioritizes the optimization of the model training process based on considering aforementioned improvement. CT-GINDTI represents drugs as graphs and utilizes graph isomorphism network to better capture the inherent structural and relational properties of drugs. Additionally, we introduce a cyclic training method to address the imbalance issue between positive and negative samples by selecting more reliable negative samples. To evaluate the performance of CT-GINDTI, we conducted extensive experiments and compared its results with seven state-of-the-art methods in the field. The experimental results demonstrate that our proposed CT-GINDTI outperforms these existing methods, showcasing its superior achievement in the prediction of DTIs.

1. Introduction

Target refers to biological macromolecules that have pharmacological functions in the body and can interact with drugs to achieve therapeutic purposes, such as certain proteins and nucleic acids (Deans et al., 2016). The occurrence of diseases is generally associated with the abnormal expression of specific targets in the human body, and drugs can impact the disease process by regulating the activity of targets. Actually, the difficulty of drug discovery lies in the limited number of known drug–target interactions (DTIs) (Xia et al., 2010). Prediction of DTIs has thus become a crucial step in the drug discovery or repositioning process (Luo et al., 2017). The prediction of DTIs can be performed using the experimental (in-vivo) methods, however, it is widely acknowledged that experimental determination of DTIs is extremely costly, time-consuming, and limited to small-scale research (Haggarty et al., 2003). In contrast, computational approaches can significantly reduce the range of drug candidates for downstream experimental validation, thereby greatly reducing the high cost and the long period of developing a new drug (Wan et al., 2019).

There is a growing number of researchers are using machine learning algorithms to predict DTIs over recent years. Based on the assumption that drugs with similar structures are more likely to have similar targets (Wang et al., 2021), DTI prediction is usually formulated as a binary classification task aimed at predicting the probability of DTI existence (Redkar et al., 2020; Johnson et al., 1990). In 2008, Yamanishi et al. proposed the concept of "pharmacological space" for the first time and developed a bipartite graph model that integrates the chemical and genomic spaces into the drug-protein interaction network (Yamanishi et al., 2008). Subsequently, in 2009, Bleakley and Yamanishi transformed edge-prediction problems into binary classification problems and applied a support vector machine framework to predict DTIs using a Bipartite Local Model (BLM) (Bleakley et al., 2009). However, these methods did not utilize a large amount of unlabeled information to assist in prediction. Xia et al. proposed a semi-supervised learning method called NetLapRLS, which utilizes Laplacian regularized least squares (RLS) and incorporates both similarity and interaction kernels into the prediction framework (Xia et al., 2010). To provide a reasonable prediction for drug/target candidates that are currently new, Mei et al.

* Corresponding author.
E-mail addresses: yhdu@lut.edu.cn (Y. Du), yaoyabing@lut.edu.cn (Y. Yao), tangjx@lut.edu.cn (J. Tang), zhaozhl@lzu.edu.cn (Z. Zhao), zhuoyue_gou@163.com (Z. Gou).

https://doi.org/10.1016/j.eswa.2024.123730
Received 30 October 2023; Received in revised form 23 December 2023; Accepted 17 March 2024
Available online 19 March 2024

integrated BLM with a Neighbor-based Interaction-profile Inferring (NII) procedure, which can learn DTI features from neighbors and predict interactions for new drug or target candidates (Mei et al., 2013). Additionally, van Laarhoven et al. proposed WNN-GIP, which can be directly applied to processing unknown targets, or both unknown drugs and targets (van Laarhoven et al., 2013).

Using information from multiple data sources can improve the accuracy of feature learning. Luo et al. presented DTINet, which integrates heterogeneous data sources information and constructures an informative feature matrix for learning drug and target features (Luo et al., 2017). Wan et al. developed NeoDTI, a new nonlinear end-to-end learning model that uses neural network methods to learn drug and target features based on DTINet (Wan et al., 2019). In a different approach proposed by Sun et al., AEFS, which is based on an autoencoder and predicts DTIs while considering spatial consistency constraints (Sun et al., 2021).

In the past few years, various end-to-end approaches have been proposed, with a focus on extracting protein and pharmacological properties from sequences using various neural networks (Hu et al., 2023). One notable example is DeepDTIs, developed by Wen et al., which applies a deep belief network (DBN) to extract the features of drugs and targets from chemical substructure and sequence. Zong et al. proposed a method which adopts DeepWalk to calculate the similarities within the Linked Tripartite Network (LTN) (Zong et al., 2017). Wang et al. used stacked autoencoder to extract information from drug molecular structures and protein sequences, and predicted DTIs through the rotation forest (Wang et al., 2018). DLDTI, proposed by Zhao et al., combines network representation learning and convolutional neural networks (Zhao et al., 2020). Huang et al. proposed an augmented transformer (Vaswani et al., 2017) encoder-based method for extracting and capturing semantic relations among substructures of drugs and targets from a large amount of unlabelled biological data (Huang et al., 2021).

Notably, Graph Convolutional Networks (GCN) (Kipf et al., 2017) have gained significant popularity in DTI prediction due to their ability of dealing with non-Euclidean data (Sun et al., 2022; Wen et al., 2019). Manoochehri et al. utilized encoders to create feature vectors and reconstructed edge labels using decoders (Manoochehri et al., 2020). The GADTI proposed by Liu et al., in which the encoder consists of GCN and Random Walk with Restart (RWR) to provide more information about the nodes (Liu et al., 2021), and DisMult (Yang et al., 2014) is used as the decoder. The GANDTI model proposed by Sun et al. establish a graph convolutional autoencoder to learn the network embedding, and a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) is introduced to regularize the feature vectors of nodes to satisfy a Gaussian distribution (Sun et al., 2022). Chen et al. developed SDGAE, which first uses the Weighted K Nearest Known Neighbours (WKNKN) to reduce isolated nodes and utilizes Spatial Consistency Constraint (SCC) to maintain invariant neighbor relationships during representation learning (Chen et al., 2023). Nguyen et al. proposed a new model called GraphDTA that represents drugs as graphs and employs multiple graph neural network architectures to predict Drug–Target Affinity (DTA) (Nguyen et al., 2021). Qing et al. presented a graph auto-encoder and multi-subspace deep neural network-based DTI prediction method (Ye et al., 2022).

While the aforementioned deep learning methods have shown promising performance, there are still areas for improvement. Firstly, most of these models represent drugs as strings, which results in the loss of structural information of the molecule and may impair the predictive power of the model as well as the functional relevance of the learned latent space (Nguyen et al., 2021). Secondly, the selection of real non-interacting drug-target pairs as negative samples in previous prediction tasks remains challenging, it is crucial to find a reliable method to create credible negative sample datasets for supervised learning (Bagherian et al., 2020). Lastly, it is important to evaluate the performance of deep learning methods in real drug development scenarios, as good performance on testing datasets does not necessarily guarantee success in practical applications (Bagherian et al., 2020).

To try to alleviate the above situation, in this paper, we present CT-GINDTI for DTI prediction. CT-GINDTI is a novel neural network architecture that directly models drugs as molecular graphs and employs graph isomorphism network convolutional layers to extract features from graph structures. Additionally, a reliable negative sample dataset construction method called cyclic training is introduced to reduce the possibility of mistakenly selecting unknown but interacting drug-target pairs as negative samples. Through comprehensive tests, it has been demonstrated that CT-GINDTI achieves substantial performance improvement over other state-of-the-art prediction methods.

2. Methods

2.1. Graph isomorphism network (GIN)

Learning with graph-structured data, such as molecules, social, biological, and financial networks, requires effective representation of their graph structure (Hamilton et al., 2017). Recently, there is a strong interest in Graph Neural Network (GNN) approaches for graph representation learning (Kipf et al., 2017; Tarlow et al., 2016; Xu et al., 2018). GNNs is usually a recursive neighborhood aggregation scheme, where each node aggregates its neighbor's feature vector to calculate its new feature vector (Xu et al., 2018; Gilmer et al., 2017). After k iterations of aggregation, a node is represented by its transformed feature vector, which captures the structural information within its k-hop neighborhood. The final representation of the entire graph can be obtained by pooling (Ying et al., 2018).

In 2019, Xu et al. Xu et al. (2018) proposed a structure called Graph Isomorphism Network (GIN). GIN proposes a higher but reasonable requirement, isomorphism, which means that if two graphs are isomorphic, their processed graph features should be the same, and vice versa. GIN is a powerful graph neural network model that can capture substructure information of graphs. Therefore, we utilize the GIN model for learning on the graph representation of drugs.

Formally, let $G = (V, E)$ denote a graph for a given drug, where V is the set of N nodes(C-dimensional) and E is the set of edges. A multi-layer GIN takes a node feature matrix $X \in \mathbb{R}^{N\times C}$($N$ and C are the number of atoms and the feature dimension of per node, respectively) and an adjacency matrix $A \in \mathbb{R}^{N\times N}$ as input and produces a node-level output $Z \in \mathbb{R}^{N\times F}$ (F is the number of output features per node). The GIN uses a multilayer perceptron model to aggregate feature vectors of neighbors of nodes to iteratively update its feature representation. The graph convolution operation of GIN can be represented as:

$$h_v^{(k)} = MLP^k\left(1+\epsilon^k\right)h_v^{(k-1)} + \sum_{u\in N(v)} h_u^{(k-1)} \tag{1}$$

Here, $h_v^{(k)}$ represents the feature vector of node v in layer k, $N(v)$ is the set of nodes neighboring v, ϵ is a learnable parameter or a fixed scalar used to balance the contribution between node's features and the features of its neighboring node, enabling the capture of local structural information of the graph.

2.2. Drug representation

Simplified Molecular Input Line Entry System (SMILES) (Weininger et al., 1988) is a one-dimensional representation sequence of chemical structure, which utilizes a linear sequence of letters and numbers to describe the structural features of compounds. From the SMILES, drug features used for DTI prediction can be inferred, the majority of models represent drugs as strings as drug input (Veselinović et al., 2015; Öztürk et al., 2016; Yang et al., 2021; Öztürk et al., 2018; Huang et al., 2021).

Similar to Chaudhari et al., 2021; Hu et al., 2023; Nguyen et al., 2021; Wang et al., 2022, we view drug compound as a graph of the

interactions between atoms to encompass more features of the drug. To describe nodes in the graph, we use a set of atomic features adapted from DeepChem (Ramsundar et al., 2019), which includes five pieces of information: the atom symbol, the number of adjacent atoms, the number of adjacent hydrogens, the implicit value of the atom, and whether the atom is in an aromatic structure. Each node is represented by a 78-dimensional binary feature vector, with the first 44 dimensions representing the atom symbol, the next 11 dimensions indicating the number of adjacent atoms, followed by 11 dimensions for the number of adjacent hydrogens, 11 dimensions for the implicit value of the atom, and the last dimension indicating whether the atom is in an aromatic structure. We convert the SMILES code of a drug into its corresponding molecular graph and extract atomic features using the open-source chemical informatics software RDKit (Landrum et al., 2006). Fig. 1 showcases the node feature vector of the first carbon atom in the structural diagram of Valsartan.

With drug compounds represented as graphs, the next step is to design an algorithm that effectively learns from this graphical data. We utilize the GIN network for learning the representation of drugs, where the input is node feature matrix X of all atoms in each molecular graph and adjacency matrix A of molecular graph and the output is node features. In CT-GINDTI, the graph neural network for drug features extraction consists of four GIN layers, with a batch normalization layer between GIN layers. The GIN layers iteratively update the feature vectors of the nodes in the graph based on the features of their neighbors, which allows the network to capture the structural information and relationships between atoms in the drug compounds. Finally, a global max pooling layer is used to aggregate the features from all nodes in the graph to obtain the representation of the entire graph Z. When pooling, the maximum value on each feature dimension is selected as the value of all nodes in the graph, resulting in a fixed-length feature vector that captures the most salient information from the graph structure. The detailed process of learning drug feature vectors through GIN convolutional layers in CT-GINDTI is illustrated in Fig. 2.

2.3. Protein representation

While proteins can also be represented as graphs, Here, we decided to abandon the graph representation due to the lack of reliable tertiary structure information. Instead, we represent proteins using their amino acid sequence, which is not only fundamental in determining the secondary and tertiary structures, or 'fold', but also determines the biological activity of protein.

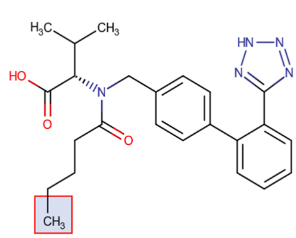
Each amino acid type in amino acid sequence is encoded with an integer based on its associated alphabetical symbol, for example, Alanine(A) is encoded as 1, Cysteine(C) as 3, Aspartic Acid(D) as 4, and so on. This allows us to represent the protein as an integer sequence. The correspondence between amino acid types and integers is shown in Table 1. To make it convenient for training, the sequence is cut or padded to a fixed length sequence of 1000 residues. In case a sequence is shorter, it is padded with zero values.

These integer sequences are then used as input to embedding layer, which outputs a 128-dimensional vector for each amino acid. Subsequently, three 1D convolutional layers are employed to learn different levels of potential features from the input sequence. Finally, a max pooling layer is applied to obtain a representation vector of the input protein sequence.

2.4. Framework of CT-GINDTI

From the *Introduction*, it is evident that existing methods for predicting DTIs have undergone enhancements from two perspectives: refining the accuracy of drug and target representations and improving prediction algorithm performance. In this paper, we present a novel deep-learning model for DTIs prediction named CT-GINDTI, which prioritizes the optimization of the model training process based on considering aforementioned improvement. CT-GINDTI takes drugs, targets, and their known interactions as input, and produces the probability of interaction between drug-target pairs as output. To elaborate, CT-GINDTI comprises two main stages: representation learning stage and cyclic training stage.

During the representation learning stage, we employ molecular graphs to represent drugs, enabling the model to directly capture the bonds among atoms. The network begins with a GIN layer that takes the node feature matrix (composed of representations of each atom in the molecular graph) and adjacency matrix (topologic information of molecular graph) as inputs, then passes a convolved feature matrix to the subsequent GIN layer and final graph representation vector is computed by the global max pooling layers from the GCN layer output. As for proteins, their amino acid sequences serve as the initial input and are categorically encoded. An embedding layer is added to the sequence, assigning a 128-dimensional vector to each character. Next, three 1D Convolutional Neural Networks (CNN) layers and a global max pooling layer are added to learn a sequence representation vector. And then, a four-layer full connect neural network is constructed and trained using the feature vectors, the resulting output value represents the probability of interaction between the drug-target pair. Please refer to Fig. 3 for an overview of the entire workflow.



Valsartan

CCCCC(=O)N(CC1=CC=C(C=C1)C1=CC=CC=C1C1=NNN=N1)[C@@H](C(C)C)C(O)=O

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. — the atom symbol: C

0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. — the number of adjacent atoms: 1

0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. — the number of adjacent hydrogens: 3

0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. — the implicit value of the atom: 3

0.] — whether the atom is in an aromatic structure: 0(no)

Fig. 1. The node feature vector of the first carbon atom in the structural diagram of Valsartan.

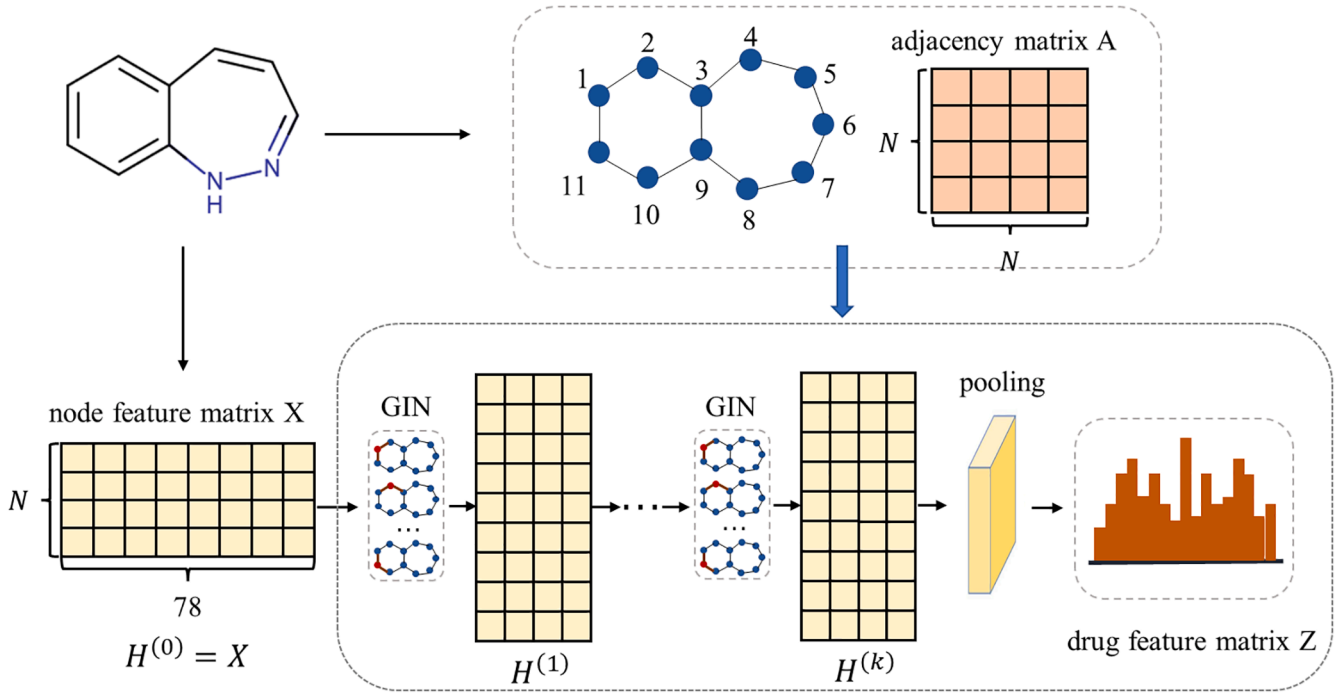


Fig. 2. Drug feature vector learning process based on GIN network.

Table 1
The correspondence between amino acid types and integers.

Name	Letter	Integers code
Alanine	A	1
Cysteine	C	3
Aspartic	D	4
Glutamic	E	5
Phenylalanine	F	6
Glycine	G	7
Histidine	H	8
Isoleucine	I	9
Lysine	K	11
Leucine	L	12
Methionine	M	13
Asparagine	N	14
Proline	P	16
Glutamine	Q	17
Arginine	R	18
Serine	S	19
Threonine	T	20
Valine	V	22
Tryptophan	W	23
Tyrosine	Y	25

Moving on to the cyclic training stage, our objective is to select actual non-interacting drug-target pairs and generate more reliable negative sample datasets. Initially, similar to most existing methods, we select known DTIs as positive examples and randomly choose an equal number of unknown DTIs as negative samples to train the CT-GINDTI model. Subsequently, the trained model is employed to predict the probabilities of all unknown drug-target pairs interacting. The prediction results are then sorted in descending order, and randomly select drug-target pairs from the last 50 % are serving as negative samples. Compared to the original dataset, non-interacting drug-target pairs are more likely to be ranked towards the end of the sequence. These negative samples are combined with an equal number of positive samples to form a current dataset, which is used to retrain the model, yielding new prediction results. Repeat the previous step until optimal performance is achieved. A schematic diagram of the cyclic training process is shown in Fig. 4.

To illustrate the reliability of selecting negative samples for cyclic training methods, we conducted the following theoretical analysis.

Assuming the total number of samples be N, where the number of positive and negative samples are m and n ($n \gg m$), respectively. select k samples as negative samples.

(1) If negative samples are randomly selected, the misselection rate P_1 will be:

$$P_1 \approx \frac{k}{N} \tag{2}$$

(2) After a round of iteration, predict the probabilities of all unknown drug-target pairs interacting, calculate the AUROC value (the probability that any positive sample has a higher score than any negative sample), and randomly select k negative samples from the last 50 % of the predicted probabilities in descending order. The calculation of the misselection rate at this time is a NP-hard problem, as the distribution of positive samples is unknown.

According to the definition of AUROC, if the AUROC value is c and the total sample size is N, then

$$\sum_{i=1}^{cN} (cN - i)P_i = \sum_{j=cN+1}^{N} (j - cN)P_j \tag{3}$$

where P_i and P_j represent the probabilities that the i_{th} and j_{th} samples in the sequence are positive samples, respectively. Assuming $P_i = \overline{P_i}$, $P_j = \overline{P_j}$, the above equation can be simplified to

$$\frac{cN(cN-1)}{2}\overline{P_i} = \frac{(N-cN)(N-cN+1)}{2}\overline{P_j} \tag{4}$$

Furthermore, given that:

$$cN\overline{P_i} + (N - cN)\overline{P_j} = 1 \tag{5}$$

By combining the above two equations, we obtain:

$$\overline{P_i} = \frac{1-c}{cN} \tag{6}$$

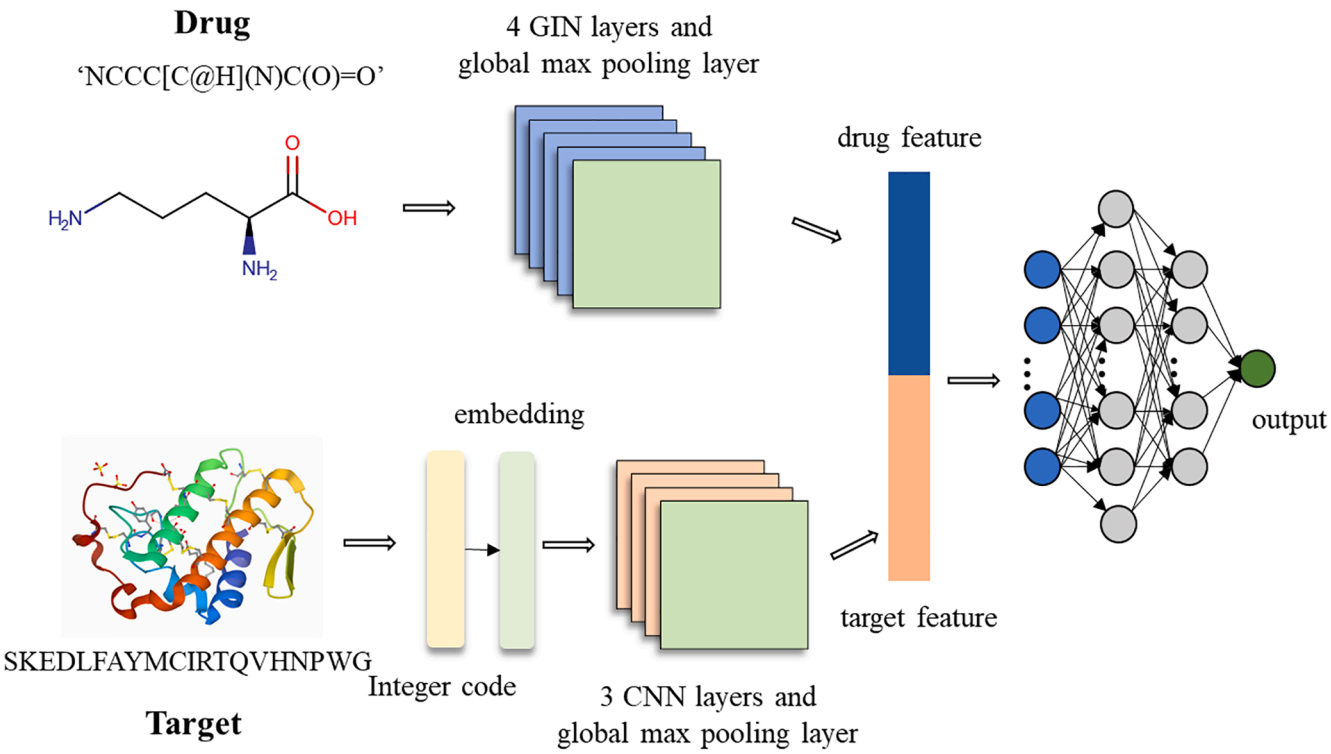


Fig. 3. The overall workflow of the representation learning stage.

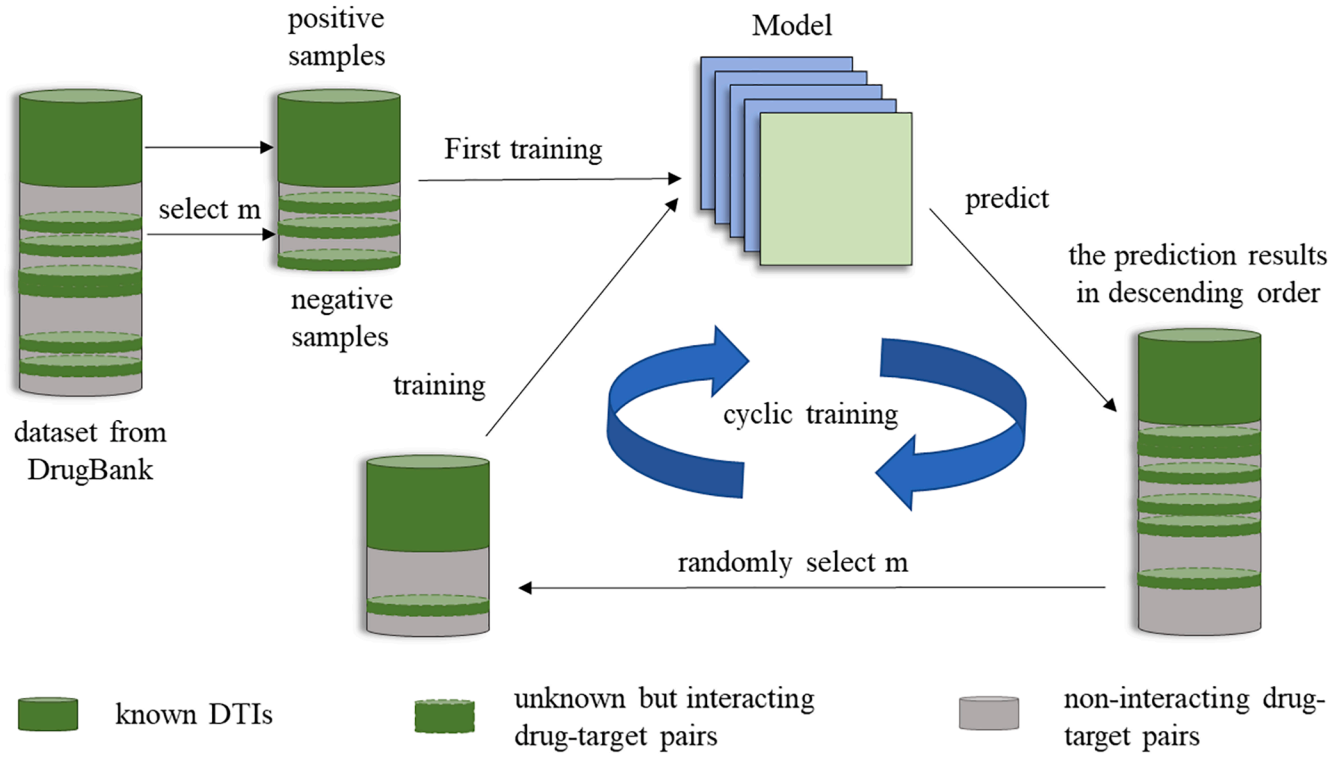


Fig. 4. Schematic diagram of cyclic training process.

the misselection rate P_2 will be:

$$P_2 \approx \frac{1-c}{cN}k \tag{7}$$

Therefore,

$$\frac{P_2}{P_1} = \frac{1-c}{c} = \frac{1-AUROC}{AUROC} \tag{8}$$

Through the theoretical proof above, it can be approximately considered that the cyclic training method would reduce the sample

misselection rate to the original (1-AUROC)/AUROC, which means that if the AUROC metric of the prediction result is 0.9, the sample misselection rate is reduced to the 1/9 of the original, effectively solving the problem of negative sample selection. Therefore, we believe that the cyclic training method can enhance the accuracy of prediction.

3. Experiments and discussion

3.1. Dataset

The data used in this study, including drug chemical structure, target sequence and known DTIs, was sourced from the public DrugBank database (version 5.1.10, released 2021-01-03) (Wishart et al., 2017). DrugBank is a comprehensive, free-to-access, online database containing information on drugs and drug targets. Because of its broad scope, comprehensive referencing, and detailed data descriptions, DrugBank is widely used in the fields of drug development and drug repositioning.

To integrate the interaction information between drugs and targets, First, we obtained structural information of 2,715 approved small molecule drugs, amino acid sequence information of 3,061 proteins, and 15,840 known drug-target interaction information from the "structures", "target-sequences" and "protein-identifiers" modules of the Drugbank database, respectively. Secondly, based on the known DTIs, we screened out the required drugs and targets from the drug structural information and protein amino acid sequence information. The final dataset used for the experiment contains 1,958 drug, 2,503 target, 15,370 known DTIs and 4,885,504 unknown DTIs. Among the drugs, 292 drugs have only one associated target, while 944 drugs have fewer than five targets, accounting for 48.21 % of the total number of drugs.

3.2. Evaluation metrics

To evaluate the performance of our model, we carried out a 5-fold cross-validation (Wong et al., 2015) by dividing the dataset into 5 subsets. Each fold was then taken in turn as a test set, while the remaining four folds served as the training set. We focused on two aspects in choosing the evaluation metrics: One side, it is crucial to determine whether the DTIs with higher scores truly exist, as biologists often select drug-target pairs with high prediction scores for subsequent wet experiment validation. Overall predictive performance is the other important aspect of model evaluation. Based on these considerations, we utilized **Recall** and **the area under the Receiver Operating Characteristic (ROC) curve (AUROC)** (Hanley et al., 1982) to assess the model's predictive capability.

The recall rate reflects the proportion of drug-target pairs with real interactions in the top L predictions, A higher recall rate indicates a more accurate model's classification of the true sample. It is calculated using the formula:

$$Recall = \frac{L_r}{L} \tag{9}$$

where L_r denotes the number of positive samples that were predicted correctly and L is the total number of positive samples in the test set.

AUROC is employed to assess the model's predictive performance by considering both positive and negative sample classification. It measures the area under the ROC curve, where the ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR).

In summary, Recall and AUROC are used as evaluation metrics to assess the predictive performance of models, taking into account both the classification of positive samples and overall classification performance.

3.3. Parameter setting

3.3.1. Number of cyclic training

In this paper, we proposed a novel training method to select genuine non-interacting drug–target pairs and generate more credible negative sample datasets, we conducted several iterations of the experiment to determine the optimal number of cyclic training. The performance of the CT-GINDTI models with different training iterations is depicted in Fig. 5. The primary coordinate axis and secondary coordinate axis correspond to the Recall and the AUROC value, respectively. The position of each data point on the chart indicates the performance of the model after a specific training, with higher values on both axes indicating better performance.

From Fig. 5, it is observed that the performance is initially poor in the first training iteration, with Recall and AUROC results of 0.0699 and 0.8769, respectively. This can be attributed to the random selection of negative samples, which may mistakenly select unknown but interacting drug-target pairs. The Recall and AUROC results of the model improved significantly with the increasing of cyclic training times, reaching optimal Recall after 4 iterations of training and AUROC after 3 times, with 60.0 % and 2.9 % higher than the first training, respectively. Notably, when the number of training was set to 4, the AUROC result is 0.9008, which is only 0.20 % lower than the optimal value of 0.9026 but the recall rate reached its maximum of 0.1120, surpassing the suboptimal value of 0.0940 by 19.03 %. Therefore, the number of cyclic training was set to 4 in subsequent experiments by considering the comprehensive analysis of both Recall and AUROC results.

Ideally, the model's performance and accuracy should continue to improve and finally tended to be stable with the number of cyclic training increases. However, our experimental results showed a fluctuation phenomenon after 4 iterations of training. Through in-depth analysis, we found that this phenomenon is mainly attributed to the following two reasons: Firstly, through theoretical analysis, we found that the cyclic training method can reduce the probability of misselection negative samples to about 1/9 of the original, which has significantly reduced misselection probability. However, due to the large overall sample size, there is still a high possibility of misselection unknown samples as negative samples, which negatively affects the overall model performance. Secondly, negative samples are randomly choosing from the last 50 % of the samples after sorting, and this randomness can lead to fluctuations in model performance to a certain extent.

To eliminate the impact of randomness on the model, when initially designing the model, we selected a portion of the lowest probability data in each prediction as negative samples. However, the model prediction effect is not satisfactory. The reason is that there were a large number of similar drug or target clusters in the negative samples, and these drugs (targets) have low activity and only interact with a small number of

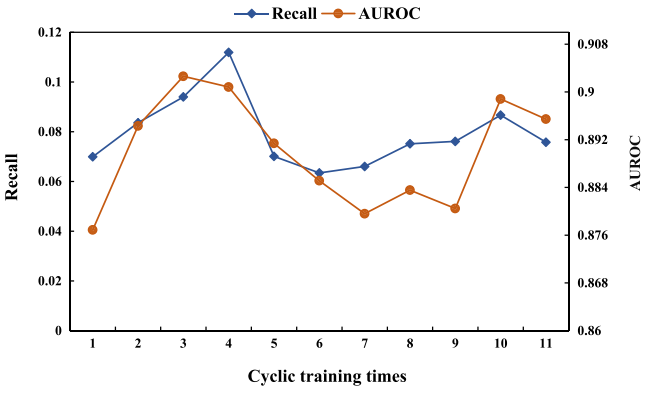


Fig. 5. The performance of the CT-GINDTI models with different training iterations.

targets (drugs). After training, the model would judge that the probability of these drugs (targets) interacting with other targets (drugs) is very low. Selecting negative samples based on the data with the lowest prediction probability would result in a large number of such drugs be chosen as negative samples, making the selection of negative samples lack diversity and reducing the prediction accuracy of the model. Therefore, we randomly select some data from the last 50 % of the prediction results as negative samples, ensuring a low false selection rate of negative samples while taking into account the diversity of sample selection. However, such a selection method inevitably leads to fluctuations in the model prediction results.

3.3.2. Optimization of architecture

For architecture optimization, all the combinations of three to seven GIN layers and 16, 32, 64, 128, and 256 nodes in each GIN layer were individually tested to evaluate the performance of CT-GINDTI and select a model with optimal architecture. Recall and AUROC results for CT-GINDTI with different architectures are shown in Table 2, the best results are highlighted in bold font, while the second best performance is indicated by an asterisk.

From Table 2, it is evident that CT-GINDTI achieved the best performance with either four or five GIN layers and 128 nodes in each GIN layer. The optimal Recall rate of 0.1119 was achieved when using five GIN layers with 128 (or 32) nodes per layer, and the second best of 0.1103 was obtained with four GIN layers and 128 nodes in each layer. In terms of AUROC, the optimal and suboptimal results were relatively close, with values of 0.9140 and 0.9111, respectively. The former was achieved with four GIN layers and 128 nodes in each layer, while the latter was obtained with three GIN layers and 64 nodes.

To visually compare the performance of CT-GINDTI with different architectures, we generated a histogram (Fig. 6), where the height of the bars corresponds to the performance. The histogram confirms the same conclusions as presented in Table 2. Taking into consideration both Recall and AUROC results, we selected CT-GINDTI with four GIN layers and 128 nodes in each layer as the optimal architecture for subsequent comparative experiments. By carefully considering the performance of CT-GINDTI with different architectures, we aim to identify the optimal model's architecture and enhance its predictive capabilities.

Moreover, the activation function in the output layer of CT-GINDTI is ReLu. Batch normalization (BN) with a size of 512 was implemented to normalize layer inputs to reduce the effects of internal covariate shift (Ioffe et al., 2015) of network. The loss function employed is Mean Square Error (MSE) and Adam optimizer was used to train our model. All the settings and hyperparameters of CT-GINDTI are summarized in Table 3.

3.4. Comparative experiment

To evaluate the validity and accuracy of our CT-GINDTI model, we compared it with four other baseline prediction methods, namely WNN-GIP, BLMNII, MSCMF and NetLapRLS. In addition, a deep learning model, GCN, and two new algorithms proposed in 2023 for DTI prediction, SDGAE and CPI-NXTFusion, were also used in the comparative experiment.

3.4.1. Brief introduction to algorithms

(1) **GCN**

The GCN model is also used to learn the graph representation of drugs. The GCN-based model takes the node feature matrix X and the adjacency matrix A of the drug molecule graph as inputs, and uses three GCN layers to learn node-level features Z. All layers use Relu as the activation function. After the last GCN layer, a max-pooling layer is added to obtain the representation vector of the entire graph. However, the cyclic training method has not been applied on GCN model. The learning process of the protein feature vector is completely consistent with that of the CT-GINDTI model. All the settings and hyperparameters of GCN model are summarized in following Table 4.

(2) **WrapperDTI**

The first method, which we refer to as WrapperDTI in the rest of the paper, consists of three phases: feature extraction, feature selection, and classification.

In feature extraction phase, various features from protein sequences have been extracted. These features include EAAC, EGAAC, DDE, TF-IDF, k-gram, BINA, PSSM, NUM, PsePSSM, PseAAC. The descriptors of the structure of drug molecular of features have used the 881-dimensional binary vector format.

In the feature selection stage, the IWSSR hybrid algorithm is applied to search for effective features in the space of features. The SU standard is applied to weight features, and it is computed as follows:

$$SU_{i,c}(F_i, C) = 2\frac{H(F_i) - H(F_i|C)}{H(F_i) - H(C)} \tag{10}$$

The selected features are then given to rotation forest classification, to have a more efficient prediction.

At the time of prediction, using the sample ω, d_{ij} in (x R_i^a) is considered as a probability that predicts whether ω belongs to λ_j or not by using the T_i classifier. Then the level of trust in the class is calculated using the average combination, the formula of which is as follows:

$$\lambda_j(\omega) = \frac{1}{L}\sum_{i=1}^{L} d_{ij}(xR_i^a) \tag{11}$$

The category with the highest probability will be considered as a test sample x.

Based on the optimal feature combination determined by Mesrabadi et al., we selected the combination of PSSM, EGAAC, EAAC, DDE, and BINA, with a feature vector length of 4625 for comparative experiments.

(3) **CPI-NXTFusion-G***

The second algorithm referred to as CPI-NXTFusion, was proposed by Mazzone et al. is a new approach based on data fusion for DTI prediction built on top of the NXTfusion library, which generalizes the Matrix Factorization (MF) paradigm by extending it to the nonlinear inference over Entity–Relation (ER) graphs.

In the classical MF data fusion paradigm, a target matrix (relation) Y

Table 2
Recall and AUROC results of CT-GINDTI with various architectures.

	16		32		64		128		256	
	Recall	AUROC	Recall	AUROC	Recall	AUROC	Recall	AUROC	Recall	AUROC
3	0.1008	0.8974	0.0683	0.9005	0.1047	0.9111*	0.0976	0.9101	0.0921	0.8985
4	0.0514	0.8707	0.0560	0.8998	0.0680	0.8969	0.1103*	**0.9140**	0.0618	0.9088
5	0.0560	0.8865	**0.1119**	0.9008	0.0569	0.8995	**0.1119**	0.9087	0.0654	0.9025
6	0.0722	0.8801	0.0553	0.8939	0.0605	0.9031	0.0579	0.9043	0.0917	0.9032
7	0.0371	0.8790	0.0449	0.8874	0.0683	0.9019	0.0618	0.8952	0.0621	0.8963

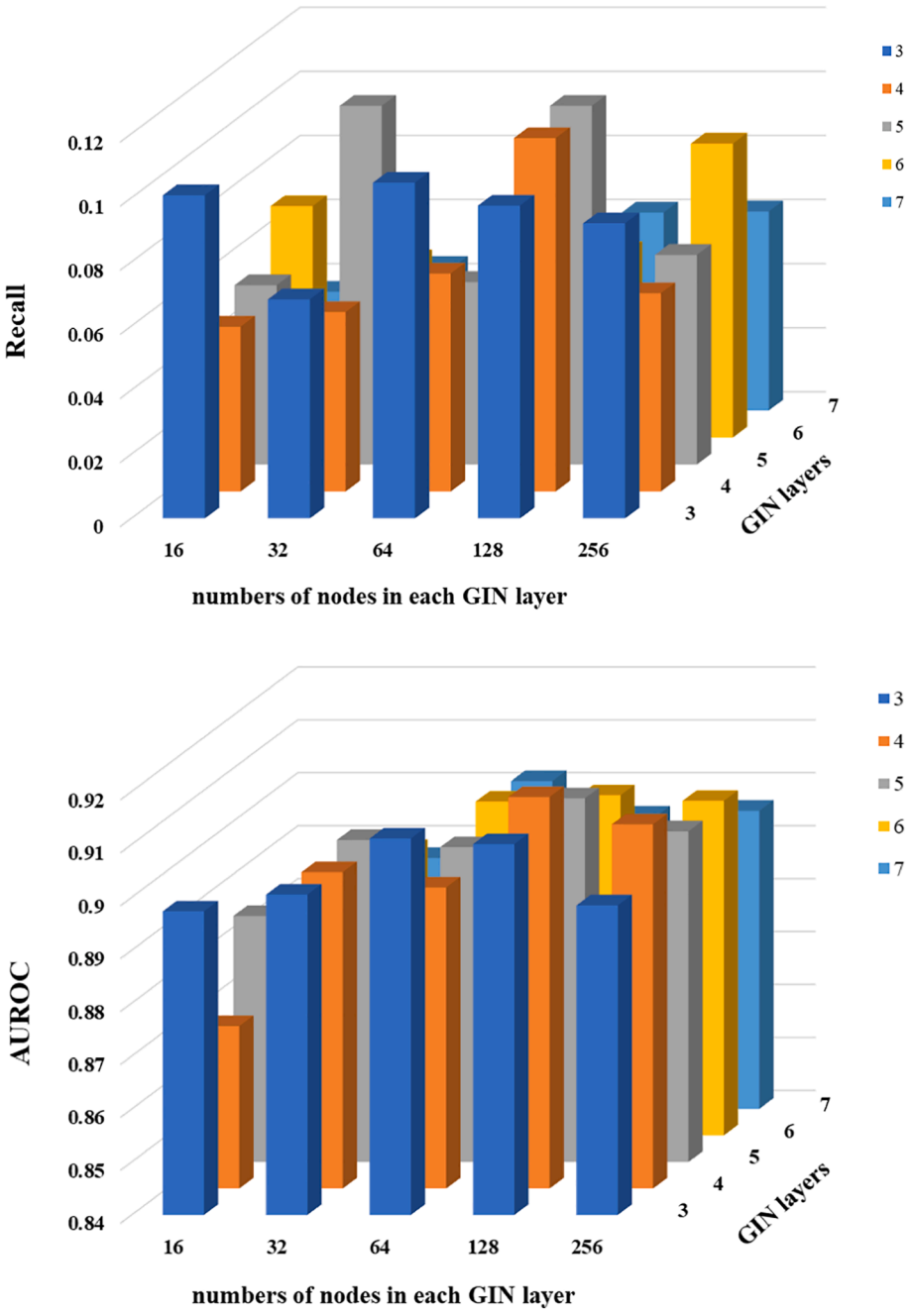


Fig. 6. Performance of CT-GINDTI with various architectures.

Table 3
Hyperparameters used in the CT-GINDTI framework.

Hyperparameter	Setting
Number of cyclic training	4
Activation function	Rectified linear unit
Batch normalization	Yes
Batch size	512
Learning rate	0.0001
Optimizer	Adam
Pooling	Global max pooling

Table 4
Hyperparameters used in the GCN model.

Hyperparameter	Setting
Activation function	Rectified linear unit
Batch normalization	Yes
Batch size	512
Learning rate	0.0001
Optimizer	Adam
Pooling	Global max pooling

$= UV$ is reconstructed by the product of two rectangular matrices U, V, such that

$$\underset{U,V}{\operatorname{argmin}} \|Y - UV\|_F + \lambda(||U||_F + \|V\|_F) \tag{12}$$

NXTFusion extends this classical MF paradigm, in which a single matrix is factorized, allowing (1) a nonlinear relationship between U and V and (2) the concurrent factorization of an arbitrary number of matrices (now called Relations) between an arbitrary number of pairs Entities.

The ER model is globally optimized to minimize, for each R_{ij}, a relation-specific loss:

$$\mathscr{L}_{ij} = L_{ij}(R_{ij}, M_{ij}(f_i(e_i), f_j(e_j))) \tag{13}$$

The final objective function to minimize with respect to the NN parameters for a given ER graph $\mathscr{G}$ is thus

$$\sum_{\forall R_{ij} \in \mathscr{G}} \omega_{ij} \mathscr{L}_{ij} \tag{14}$$

which allows for all the relations (matrices) to be learned concurrently, weighted by the task-specific scale factor ω_{ij}. The fact that each relation R_{ij} has a specific loss L_{ij} allows the NXTfusion data fusion framework to be flexible in incorporating any kind of relation, independently from the type of prediction problem it presents.

We have chosen to add the G* model of protein–protein similarity, drug–drug similarity, and Pfam domains features for target proteins for comparative experiments.

3.4.2. Experiment results and analysis

The results of the model evaluation are presented in Table 5. From the results, it is evident that the CT-GINDTI model outperforms the other models and exhibits a significant improvement over the second best method. Specifically, CT-GINDTI achieved a 19.4 % increase in Recall and a 4.1 % increase in AUROC. These experimental findings indicate that, overall, graphical models such as GCN and CT-GINDTI outperform traditional machine learning methods. This can be attributed to the fact that a graph structure could yield a better representation of drugs, capturing the bonds between atoms in drug molecules directly. In addition, CT-GINDTI further considers the isomorphism between drug molecular graphs and utilizes a cyclic training method to select more reliable negative sample, this enables the model to learn more accurate features from the input data and exhibit superior performance compared to the GCN model.

Compared with the two latest methods, WrapperDTI and NXTFusion-G*, CT-GINDTI still exhibits excellent performance. The reason is that the WrapperDTI method combineds extracted features from proteins and applied one of the wrapper feature selection methods named IWSSR to select protein features, but it uses molecular descriptors as drugs features without further utilizing drug graph structural information. The advantage of the NXTFusion-G* method lies in its ability to integrate information from heterogeneous data sources in drug, biology, and chemistry, and has good generalization ability. However, to ensure the consistency of experimental data, all comparative experiments use the dataset introduced in Section 3.1, without adding additional heterogeneous data from drugs and targets, which may be one of the reasons why the NXTFusion-G* method does not perform as well as expected.

3.5. Ablation experiment

We conduct a series of ablation studies by considering different modular combinations to further illustrate that every module of our model is indispensable. Here, we adopt the dataset proposed in Section 3.1 to conduct an ablation experiment. The following variants of CT-GINDTI are taken into account: (1) Without drug representation module based on GIN: we removed the drug representation module based on GIN and used drug-drug similarity instead. (2) Without protein representation module: we removed the protein representation module and used protein–protein similarity instead. (3) Without drug representation module based on GIN and protein representation module: we removed both the drug representation module based on GIN and protein representation module. Used drug-drug similarity and protein–protein similarity to represent drugs and targets respectively. (4) Without cyclic training method during training, keeping the drug and target representation modules unchanged.

Fig. 7 shows the results of the performance of CT-GINDTI and four variants under Recall and AUROC metrics. When a specific component of CT-GINDTI is removed, we can see that all performance metrics fall. In other words, these parts are fundamental to the model. To be specific, when both drug representation and target representation are removed, our model decreases most, as the Recall and AUROC decreased by 57.6 %, and 11.4 %. Can be seen that three designs in our model all enhance our model, and drug representation contributes the most to CT-GINDTI.

4. Conclusion

This paper proposed a novel method, CT-GINDTI, for DTI prediction. CT-GINDTI represents drugs as molecular graphs during the representation learning stage, allowing the model to directly capture the bonds between atoms. It utilizes GIN convolutional layers to abstract the molecular graph of each drug into a feature vector of latent variables. Additionally, the amino acid sequence of protein is encoded as an integer code and features are extracted through convolutional layers. Through a comprehensive analysis of existing DTI prediction methods, it is observed that they often lack in selecting reliable negative samples. To address this issue, an innovative method called the cyclic training method has been innovatively proposed to reduce the possibility of mistakenly selecting unknown but interacting drug-target pairs as negative samples. To evaluate the performance of models, a dataset containing 1,958 drugs, 2,503 targets and 15,370 known DTIs was established. CT-GINDTI was compared against seven other methods using this dataset, the results demonstrate that CT-GINDTI achieved superior performance compared to the other methods.

In future work, we plan to enhance the performance of our model by representing proteins as graphs as well, one possible approach is to construct protein graphs based on their 3D structures. This would enable us to capture the intricate spatial relationships and interactions within the protein molecules, potentially leading to more accurate predictions. Furthermore, the cyclic training method proposed in this paper offers a new idea to address the issue of imbalance of the positive and negative samples. We believe that this method can be further explored and refined to improve the overall performance of DTI prediction models.

CRediT authorship contribution statement

Yuhong Du: Data curation, Conceptualization, Investigation, Methodology, Software, Validation, Writing – original draft. **Yabing**

Table 5
Performance of different methods on the dataset.

	WNNGIP	BLMNII	NetLapRLS	MSCMF	GCN	WrapperDTI	NXTFusion-G*	CT-GINDTI
Recall	0.0151	0.0159	0.0918	0.0634	0.0823	0.0937*	0.0785	**0.1119**
AUROC	0.6155	0.8317	0.8623	0.8210	0.8787*	0.8713	0.8559	**0.9140**

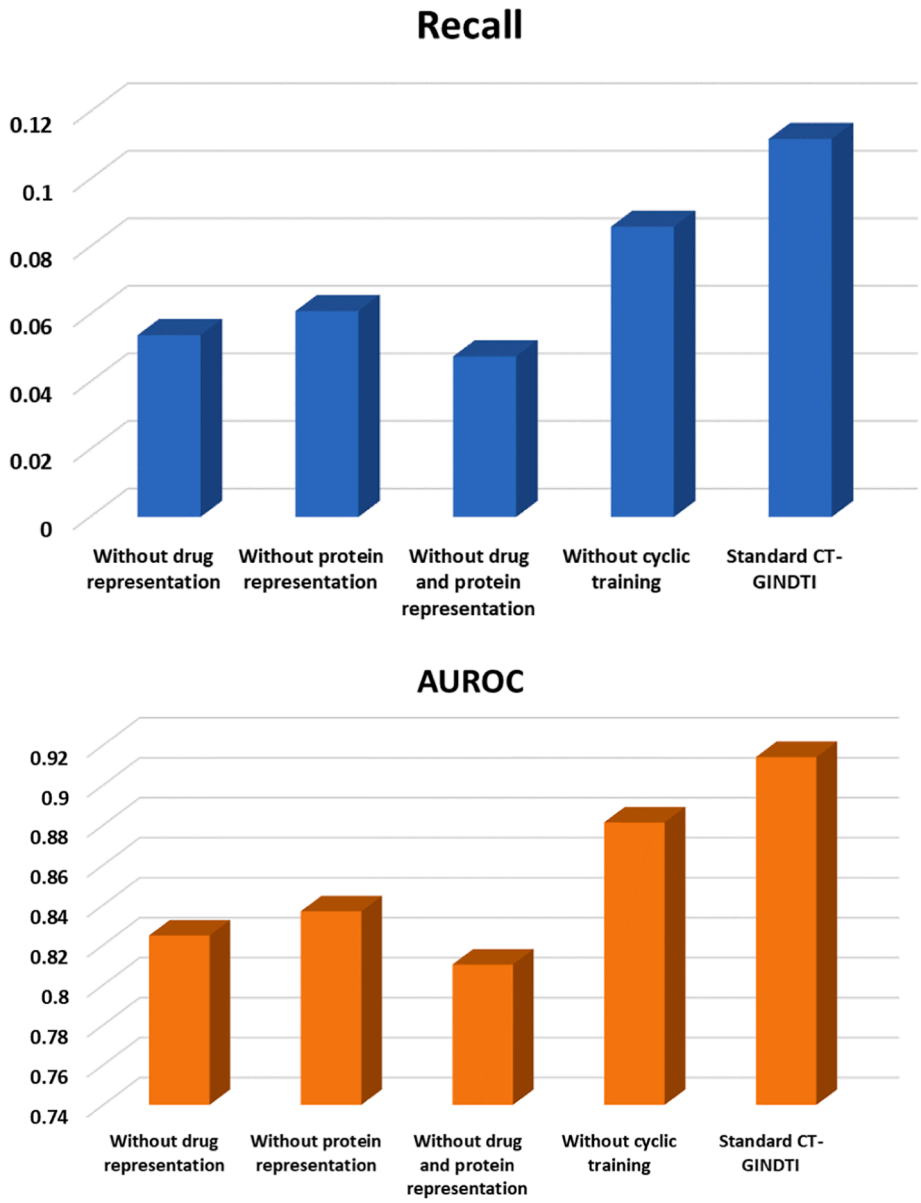


Fig. 7. The Recall and AUROC results of our ablation experiment.

Yao: Project administration, Visualization, Resources, Writing – review & editing. **Jianxin Tang:** Validation, Writing – review & editing. **Zhili Zhao:** Supervision, Writing – review & editing. **Zhuoyue Gou:** Formal analysis, Methodology, Software, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [No. 62366030]; the Gansu Provincial Natural Science Foundation [No. 23JRRA8222]; the Gansu Provincial Natural Science Foundation [No. 21JR7RA460]; the Higher Education Innovation Fund project of Gansu [No. 2022A-022].

References

Bagherian, M., Sabeti, E., Wang, K., Sartor, M. A., Nikolovska-Coleska, Z., & Najarian, K. (2020). Machine learning approaches and databases for prediction of drug–target interaction: A survey paper. *Briefings in Bioinformatics, 22*, 247–269. https://doi.org/10.1093/bib/bbz157

Bleakley, K., & Yamanishi, Y. (2009). Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*, 25(18), 2397–2403. https://doi.org/10.1093/bioinformatics/btp433.

Chaudhari, M., Thapa, N., Ismail, H., Chopade, S., Caragea, D., Köhn, M., Newman, R. H., & Kc, D. B. (2021). DTL-DephosSite: Deep transfer Learning based approach to predict dephosphorylati-on sites. *Frontiersin Cell and Developmental Biology, 9*(2021), Article 662983. https://doi.org/10.3389/fcell.2021.662983

Chen, P., & Zheng, H. (2023). Drug-target interaction prediction based on spatial consistency constraint and graph convolutional autoencoder. *BMC Bioinformatics, 24*, 151. https://doi.org/10.1186/s12859-023-05275-3

Deans, R. M., Morgens, D. W., Ökesli, A., Pillay, S., Horlbeck, M. A., Kampmann, M., Gilbert, L. A., Li, A., Mateo, R., Smith, M., Glenn, J. S., Carette, J. E., Khosla, C., & Bassik, M. C. (2016). Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. *Nature Chemical Biology, 12*(5), 361–366. https://doi.org/10.1038/nchembio.2050

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. *ICML*, 1273. https://doi.org/10.48550/arXiv.1704.01212

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *NIPS, 2*, 2672–2680. https://doi.org/10.48550/arXiv.1406.2661

Haggarty, S. J., Koeller, K. M., Wong, J. C., Grozinger, C. M., & Schreiber, S. L. (2003). Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proceedings of the National Academy of Sciences, 100* (8), 4389–4394. https://doi.org/10.1073/pnas.0430973100

Hamilton, W. L., Ying, R., Leskovec, J. (2017). Representation learning on graphs: Methods and applications. *IEEE Data Engineering Bulletin*, 40(3), 52–74. https://doi.org/10.48550/arXiv.1709.05584.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hu, J. Y., Yu, W., Pang, C., Jin, J. R., Pham, N. T., Manavalan, B., & Wei, L. Y. (2023). DrugormerDTI: Drug graphormer for drug-target interaction prediction. *Computers in Biology and Medicine, 161*, Article 106946. https://doi.org/10.1016/j.compbiomed.2023.106946

Huang, K. X., Fu, T. F., Glass, L. M., Zitnik, M., Xiao, C., & Sun, J. M. (2021). DeepPurpose: A deep learning library for drug-target interaction prediction. *Bioinformatics, 36*(22–23), 5545–5547. https://doi.org/10.1093/bioinformatics/btaa1005

Huang, K., Xiao, C., Glass, L. M., & Sun, J. (2021). MolTrans: Molecular interaction transformer for drug-target interaction prediction. *Bioinformatics, 37*(6), 830–836. https://doi.org/10.1093/bioinformatics/btaa880

Ioffe, S., Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *JMLR.org*, 2015. https://doi.org/10.48550/arXiv.1502.03167.

Johnson, M., & Maggiora, G. (1990). Concepts and applications of Molecular Similarity. *Journal of Computational, 13*(4), 539–540. https://doi.org/10.1002/jcc.540130415

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *ICLR, 4*, Article 02907. https://doi.org/10.48550/arXiv.1609.02907

Landrum, G. (2006). RDKit: Open-source cheminformatics. https://www.rdkit.org/.

Li, Y. J., Tarlow, D., Brockschmidt, M., & Zemel, R. (2016). Gated graph sequence neural networks. *ICLR, 2016*, Article 05493. https://doi.org/10.48550/arXiv.1511.05493

Liu, Z. X., Chen, Q. F., Lan, W., Pan, H. M., Hao, X. K., & Pan, S. R. (2021). GADTI: Graph autoencoder approach for DTI prediction from heterogeneous network. *Front Genet, 9*(12), Article 650821. https://doi.org/10.3389/fgene.2021.650821

Luo, Y. N., Zhao, X. B., Zhou, J. T., Yang, J. L., Zhang, Y. Q., Kuang, W. H., Peng, J., Chen, L. G., & Zeng, J. Y. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communication, 8*, 573–585. https://doi.org/10.1038/s41467-017-00680-8

Manoochehri, H. E., & Nourani, M. (2020). Drug-target interaction prediction using semi-bipartite graph model and deep learning [J/OL]. *BMC Bioinformatics, 21*(4), 248. https://doi.org/10.1186/s12859-020-3518-6

Mei, J. P., Kwoh, C. K., Yang, P., Li, X. L., & Zheng, J. (2013). Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics, 29*(2), 238–245. https://doi.org/10.1093/bioinformatics/bts670

Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: Predicting drug-target binding affinity with graph neural networks. *Bioinformatics, 37*(8), 1140–1147. https://doi.org/10.1093/bioinformatics/btaa921

Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics, 34*(17), 821–829. https://doi.org/10.1093/bioinformatics/bty593

Öztürk, H., Ozkirimli, E., & Özgür, A. (2016). A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. *BMC Bioinformatics, 17*, 128. https://doi.org/10.1186/s12859-016-0977-x

Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). *Deep learning for the life sciences: Applying deep learning to genomics, microscopy, drug discovery, and more.* Sebastopol, CA, US: O'Reilly Media.

Redkar, S., Mondal, S., Joseph, A., & Hareesha, K. S. (2020). A machine learning approach for drug-target interaction prediction using wrapper feature selection and class balancing. *Molecular Informatics, 39*(5), e1900062.

Sun, C., Cao, Y., Wei, J. M., & Liu, J. (2021). Autoencoder-based drug-target interaction prediction by preserving the consistency of chemical properties and functions of drugs. *Bioinformatics, 37*(20), 3618–3625. https://doi.org/10.1093/bioinformatics/btab384

Sun, C., Xuan, P., Zhang, T., & Ye, Y. (2022). Graph convolutional autoencoder and generative adversarial network-based method for predicting drug-target interactions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19* (1), 455–464. https://doi.org/10.1109/TCBB.2020.2999084

van Laarhoven, L., & Marchiori, E. (2013). Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS ONE, 8*(6), e66952.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *NIPS, 17*(30), 6000–6010. https://doi.org/10.48550/arXiv.1706.03762

Veselinović, A. M., Veselinović, J. B., Živković, J. V., & Nikolić, G. M. (2015). Application of SMILES notation based optimal descriptors in drug discovery and design. *Current Topics in Medicinal Chemistry, 15*(18), 1768–1779. https://doi.org/10.2174/1568026615666150506151533

Wan, F. P., Hong, L. X., Xiao, A., Jiang, T., & Zeng, J. Y. (2019). NeoDTI: Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. *Bioinformatics, 35*(1), 104–111. https://doi.org/10.1093/bioinformatics/bty543

Wang, A., & Wang, M. (2021). Drug-target interaction prediction via dual Laplacian graph regularized matrix completion. *Biomed Research International, 26*(2021), 5599263. https://doi.org/10.1155/2021/5599263

Wang, L., You, Z. H., Chen, X., Xia, S. X., Liu, F., Yan, X., Zhou, Y., & Song, K. J. (2018). A computational-based method for predicting drug-target Interactions by using stacked autoencoder deep neural network. *Journal of Computational Biology, 25*(3), 361–373. https://doi.org/10.1089/cmb.2017.0135

Wang, S., & Zhao, H. (2022). SADeepcry: A deep learning framework for protein crystallization propensity prediction using self-attention and auto-encoder networks. *Briefings in Bioinformatics, 23*(5), Article bbac352. https://doi.org/10.1093/bib/bbac352

Weininger, D. (1988). SMILES: A chemical language and information system. *Journal of Chemical Information and Computer Sciences, 28*(1), 31–36. https://doi.org/10.1021/ci00057a005

Wen, T., & Altma, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *Journal of Chemical Information and Modeling, 59*(10), 4131–4149. https://doi.org/10.1101/473074

Wishart, D. S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., Wilson, M. (2017). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 2017 Nov 8. https://doi.org/10.1093/nar/gkx1037.

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross-validation. *Pattern Recognition, 48*(9), 2839–2846. https://doi.org/10.1016/j.patcog.2015.03.009

Xia, Z., Wu, L. Y., Zhou, X. B., & Wong, S. T. C. (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC System Biology, 4* (Suppl 2), S6. https://doi.org/10.1186/1752-0509-4-S2-S6

Xu, K., Hu, W. H., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks. *ICLR.* , Article 00826. https://doi.org/10.48550/arXiv.1810.00826

Xu, K., Li, C. T., Tian, Y. L., Sonobe, T., Kawarabayashi, K., & Jegelka, S. (2018). Representation learning on graphs with jumping knowledge networks. *ICML, 2018*, 5453–5462. https://doi.org/10.48550/arXiv.1806.03536

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics, 24*(13), 232–240. https://doi.org/10.1093/bioinformatics/btn162

Yang, B. S., Yih, S. W., He, X. D., Gao, J. F., & Deng, L. (2014). Embedding entities and relations for learning and inference in knowledge bases. *ICLR, 1412*, Article 65752768038. https://doi.org/10.48550/arXiv.1412.6575

Yang, Z. D., Zhong, W. H., Zhao, L., & Chen, C. Y. (2021). ML-DTI: Mutual learning mechanism for interpretable drug-target interaction prediction. *Journal of Physical Chemistry Letters, 12*(17), 4247–4261. https://doi.org/10.1021/acs.jpclett.1c00867

Ye, Q., Zhang, X. L., & Lin, X. L. (2022). Drug-target interaction prediction via graph auto-encoder and multi-subspace deep neural networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20*(5), 2647–2658. https://doi.org/10.1109/TCBB.2022.3206907

Ying, R., You, J. X., Morris, C., Ren, X., Hamilton, W. L., & Leskovec, J. (2018). Hierarchical graph representation learning with differentiable pooling. *NIPS.* , Article 08804. https://doi.org/10.48550/arXiv.1806.08804

Zhao, Y., Zheng, K., Guan, B., Guo, M. M., Song, L., Gao, J., Qu, H., Wang, Y. H., Shi, D. Z., & Zhang, Y. (2020). DLDTI: A learning-based framework for drug-target interaction identification using neural networks and network representation. *Journal of Translational Medicine, 18*(1), 434. https://doi.org/10.1186/s12967-020-02602-7

Zong, N. S., Kim, H., Ngo, V., & Harismendy, O. (2017). Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics, 33*(15), 2337–2344. https://doi.org/10.1093/bioinformatics/btx160