DCDTI: Dual-Channel Neural Network for Drug-Target Interaction Prediction

Ping Zhang[1], Yongbin Zeng[1, 2, *]

[1]School of Computer, Baoji University of Arts and Sciences, Baoji 721016, China

[2]School of Information Sciences & Engineering, Lanzhou University, Lanzhou 730000, China

*Corresponding email: zengyb2025@lzu.edu.cn

Abstract

Background: The computational identification of drug-target interaction (DTI) is pivotal in drug discovery and chemical genomics. Current network-based approaches model DTI as a link prediction problem utilizing bipartite graphs. However, simplistic representations fail to encapsulate crucial biological semantic information, and how to effectively integrate molecular structure features with graph neural networks has emerged as a non-negligible challenge in the DTI domain. Results: To address these issues, we propose a Dual-Channel Neural Network for Drug-Target Interaction (DCDTI) prediction model based on GLTransformer, aiming to learn high-quality representations of the topological structures of drugs and targets with precision. Extensive experiments validate that DCDTI surpasses state-of-the-art methods. Case studies have further confirmed its generalization ability in actual DTI scenarios. Conclusion: DCDTI provides a powerful method for DTI prediction, which can also serve as a screening tool for studies of drug discovery.

Keywords

Drug-target interaction, Biological property semantics, Prediction, GLTransformer

Introduction

Identifying drug-target interaction (DTI) is a key step in drug discovery. Although laboratory identification in the wet lab is reliable for detecting drug-target interactions, it is subject to expensive and laborious costs. For instance, traditional molecular docking and molecular simulation methods tend to be limited by the three-dimensional structure with a small number of known ligands. With the collection of biomedical DTI databases and the advancement of deep learning, computational methods for DTI prediction have received increasing attention and can be used as alternative pre-screening tools for candidate DTI.

Computational DTI identification tasks can be mainly divided into the drug-target interaction binary classification and the drug binding affinity prediction. According to our research object, we focus on the classification task (i.e., DTI prediction) and it can be divided into three types of methods: machine learning-based methods, graph neural networks-based methods and biological sequence-based methods. Machine learning-based methods usually employ manual feature engineering to extract drug (target) features and make DTI prediction by typical classifiers such as Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost) and Random Forest [1-4]. For example, DrugE-Rank adopted ensemble learning to improve DTI prediction performance and DDR proposed a nonlinear fusion strategy to combine the similarity profile of targets, which serves as the sample feature to infer DTI via a random forest classifier [5-7]. Despite their effectiveness, they depended heavily on the prior knowledge of experts and the generalization of classifiers.

Graph-based methods (also known as network-based methods) have also been widely applied in recent years. A graph or network can represent the complex information of neighbour topology among different biological objects, such as drugs, proteins, and diseases. In a bipartite graph, the drug and the protein can be viewed as the nodes, and the known interactions between them can be regarded as the edges. Currently, graph-

based approaches dominate the tendency for the prediction of DTI [8-9]. For instance, Yang et al. designed a multi-channel graph convolutional neural network model using two independent graph attention networks to learn the interaction representation [10]. Xuan et al. constructed a three-layer heterogeneous network by combining multi-scale neighbouring topologies and cross-modal similarities, followed by an auto-encoder to generate low-dimensional feature representations [11]. Zhao et al. model the drug-protein pairs as a bipartite graph and utilize Graph Convolutional Network (GCN) to learn the latent feature representations of drug-protein pairs [12]. Though they widely employ the GCN to extract features from networks, the issue of over-smoothing that originated from the GCN is not well resolved. For example, molecular fingerprints and simplified molecular input line entry system (SMILES) are encoded based on molecular structure, so some molecular details, such as stereoisomerism, charge distribution, and solvent effects, are lost. Additionally, these methods fail to effectively capture the crucial information required for DTI prediction, namely the docking poses and spatial three-dimensional (3D) structures of the drug (targets). Despite their acceptable performance in DPI prediction, they are still limited due to the absence of spatial three-dimensional structures, which affect the stability and accuracy of the model.

Biological sequence-based methods aim to capture the potential local features of sequences from drugs and proteins as the sample feature to conduct DTI prediction. For instance, Gao et al. proposed an interpretable architecture, in which the features of drugs and targets were extracted by long short-term memory (LSTM) and graph convolutional network, respectively. Only using sequence information as input to predict drug-target binding affinities, Deep Drug-Target binding Affinity (DeepDTA) prediction utilized a convolutional neural network (CNN) to learn the potential representation of drug and target [13]. Similar to the work of Öztürk et al, TransformerCPI used sequence information to input the transformer framework to acquire drug-target embedding information [14]. Although biological sequence-based methods have promoted DTI identification to some extent, their performance is still limited due to the neglect of the global information of the drug-target mutual atomic bond structure semantics. It should be stressed here that global information is inclined to capture the impact of targets on drugs or drugs on targets, such that this strategy can obtain better feature representations. We believe that it is natural and reasonable to consider the impact of the efficacy of global information when learning to represent the drugs and targets for DTI prediction. Regretfully, biological sequence-based methods failed to notice this.

To address the above issues, we design a drug-target interaction prediction model based on GLTransformer called DCDTI, which contains three main modules by combining local attributes and global molecular graphs that are rich in biochemical semantics. The multi-modal data input module has two channels. One is used to capture local sequence features, and the other is used to encode global features, thereby achieving mutual adaptation between global and local features. The multi-modal data input module also aims to extract the local and global features of drugs and targets initially. The feature extraction module deeply extracts potential feature representations of drugs and targets, and feeds the features to the GLTransformer. The GLTransformer module aims to learn the global and local interactions between drugs and targets. Meanwhile, to prevent the over-smoothing problem of Graph Neural Network (GNN) at the global level, we propose a new edge update graph convolution network (EGCN) on the basis of the graph convolutional network. This EGCN aims to comprehensively learn the potential high-dimensional representation of topological structures in drugs and targets.

In summary, DCDTI combines the advantages of local-scale attributes and global-scale molecule graphs. It allows itself to not only adaptively acquire local-scale representations of drugs and proteins, but also effectively utilize the structural information of drugs and proteins. Our contributions are summarized below.

(1) The GLTransformer architecture is specifically developed to effectively capture both global and local features of drugs and targets, thereby facilitating a more nuanced representation of their characteristics.

(2) We develop a new edge update graph convolution network (EGCN) by updating both the chemical atom

(node) embeddings and bond (edge) embeddings simultaneously. The co-updating strategy provides a new perspective to learn a comprehensive drug representation, ensuring that deep information about the drug is completely retained.

(3) We integrate the frequent closed sequences (FCS) text mining algorithm for drug SMILE sequence features into the drug initial molecular representation. Dynamic convolutional neural networks (DyCNN) is utilized to acquire protein contact map embedding representation.

Materials and methods

Data collection

DCDTI is evaluated on two benchmark datasets, named Database of Useful Decoys-Enhanced (DUD-E) and Human dataset, respectively. The DUD-E dataset consists of 22886 active compounds against 102 targets. For each active compound, 50 decoys are generated, which have similar physico-chemical properties but dissimilar 2-D topologies to the active compound. Finally, we obtain 22,645 positive examples and 37,145 negative examples. The human dataset contains highly credible positive and negative DTI samples extracted by a systematic screening framework according to a similarity rule. The dataset comprises 3,369 positive interactions and 1,052 negative examples. In addition, the numbers of positive examples and negative examples in DUD-E and the Human dataset are shown in Table 1.

Table 1. Summary of the benchmark datasets.

Datasets	Positive	Negative	Interactions
DUD-E	22,645	37,145	59,790
Human	3,369	1,052	4,421

Overview of DCDTI

Figure 1 shows the network architecture of DCDTI. It is mainly composed of a multi-modal data input module of drugs and targets, the feature extraction module of sequence and molecular structure of drugs, and targets and the GLTransformer module. The multi-modal input module separately processes drugs' SMILES and molecular graphs, as well as targets' amino acid sequences and contact maps. These heterogeneous data are then transformed into unified embeddings for subsequent extraction. These embeddings are then passed to the feature extraction module. For the feature extraction module of sequence and molecular structure of drugs and targets, to get sequence features of the drug and protein, the learned features of the drug and protein are extracted through 2-Dimensional Convolutional Layer (Conv2d). Protein sequence features and drug sequence features are fed into GLTransformer's encoder and decoder, respectively. To get the protein's contact map features, DyCNN is used to extract the contact map into a fixed feature matrix. To get the drug's molecule graph features, EGCN is used to extract the drug's molecule graph features. Then, the protein's contact map features and the drug's molecule graph features are fed into GLTransformer's decoder and encoder, respectively. The GLTransformer module consists of more encoders and decoders. The details of encoder and decoder are shown in (C) and (D).

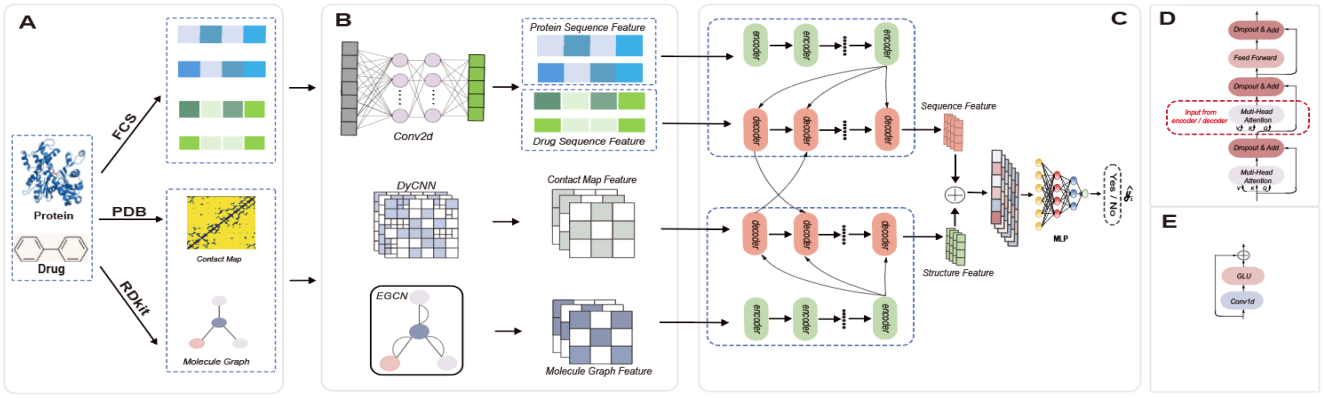

Figure 1. Overview of the DCDTI. (A) Multi-modal data input module of drugs and targets. (B) The feature extraction module of sequence and molecular structure of drugs and targets. (C) GLTransformer's struct and DCDTI's fully connected neural network. (D) The network structures of the decoder. (E) The network structures of the encoder.

Multi-modal data input module

(1) Local features

The local features representation learning is outlined as follows. Initially, the FCS text mining algorithm is utilized to decompose the SMILES strings of drugs and the protein's amino acid sequences into a collection of sub-structures.

In the FCS text mining algorithm, we first establish an initial vocabulary set, which comprises unique drugs' SMILES strings and proteins' amino acid sequences. Subsequently, a tokenized corpus containing drug (protein) tokens is created. Utilizing this tokenized corpus, we construct a feature vector with a length equal to the maximum threshold $\rho = 64$. The value at each position reflects the frequency of occurrence of the corresponding drug (protein) substructure within the sequence string.

In this way, we can get the drug features $D_{drug} \in R^{d_{length} \times \rho}$ and target features $P_{target} \in R^{P_{length} \times \rho}$, where d_{length} and P_{length} are the number of drugs and targets, respectively.

(2) Global features

For drugs, we can naturally represent the drug set as $\{G = (V_n, E_n)\}$, where $V_n \in R^{N_n \times f}$ and $E_n \in R^{N_n \times N_n \times C_{edge}}$ are the node feature matrix and adjacency matrix containing edge information of the nth drug. N_n is the number of atoms in the nth drug. f and C_{edge} are the number of feature channels of nodes and edges, respectively. For targets, we construct the contact map matrix of proteins by using the protein's Protein Data Bank (PDB) file. For example, firstly, a contact map matrix $P \in R^{d \times d}$, where d denotes the target's number of amino acid residues.

(3) Data input

Finally, in this way, we can get the drug's local features (sequence features) $D_{drug} \in R^{d_{length} \times \rho}$, protein's local features (sequence features) $P_{target} \in R^{P_{length} \times \rho}$. Additionally, we obtain the drug's global features (molecule structure features) $\{G = (V_n, E_n)\}$ and protein's global features (contact map) $P \in R^{d \times d}$.

Feature extraction module

(1) Drug and protein local features extraction

The multi-scale two-dimensional convolution (Conv2d) is used to extract the sequence feature of targets and drugs. We can get the feature matrix of the target sequence $P_{target} \in R^{P_{length} \times \rho}$ and the drug sequence $D_{drug} \in R^{d_{length} \times \rho}$. Multiple filters with different convolution kernels are used to extract the representative features of local semantic information, and the width of the sliding window for the ith convolution kernel is k_i.

$$C_{prot} = \sigma(P_{target} * W_{ip}^{cnn} + b_{ip}^{cnn}) \quad (1)$$

$$C_{drug} = \sigma(D_{drug} * W_{id}^{cnn} + b_{id}^{cnn}) \quad (2)$$

Where $W_{ip}^{cnn} \in R^{P_{length} \times \rho}$, $W_{id}^{cnn} \in R^{P_{length} \times \rho}$, $id, ip \in \{1, \cdots, n\}$, $b_{ip}^{cnn} \in R^{\rho}$, $b_{id}^{cnn} \in R^{\rho}$ are the learnable parameters, and $\rho = 64$ is the feature dimension of hidden layers, n is the number of different convolution kernels. σ is the non-linear function, $C_{prot} \in R^{P_{length} \times \rho}$ and $C_{drug} \in R^{d_{length} \times \rho}$ are output features.

(2) Protein contact map feature extraction

We construct the contact map matrix of proteins by using the protein's PDB file. Given that the contact map sizes of each protein vary, Dynamic Neural Network (DyNN) uses a convolution operation to transform these matrices of different dimensions into a fixed dimension. For example, firstly, a contact map matrix $P \in R^{d \times d}$ is transformed into a feature map matrix $P_c \in R^{d \times f}$, where f is the number of filters of the residual block. Second, we obtain an intermediate weight matrix $A_p \epsilon R^{r \times d}$, which is derived by a two-layer perceptron without bias from P_c:

$$A_p = softmax\left(W_{p2} tanh\left(W_{p1} P_c^T\right)\right) \quad (3)$$

where W_{p1} and W_{p2} are learnable parameters, r is the assumed parameter. Finally, we can get the target feature map $A_a \epsilon R^{r \times f}$:

$$A_a = A_p \times P_c \quad (4)$$

Drug molecule graph feature extraction

For drugs, we can naturally represent the drug set as $\{G = (V_n, E_n)\}$ where $V_n \in R^{N_n \times f}$ and $E_n \in R^{N_n \times N_n \times C_{edge}}$ are the feature matrix and adjacency matrix containing edge information of the nth drug. N_n is the number of atoms in the nth drug. f and C_{edge} are the number of feature channels of nodes and edges, respectively. To compensate for the lack of edge information in the GCN, we introduce a novel EGCN architecture that simultaneously processes node and edge information to comprehensively preserve the deep information in the drug graph representation. The EGCN applied to the ith drug is defined as $f(D_i)$ with a layer-wise operation as:

$V^{(l+1)} = Update\left(V^l, \delta(E_k^{(l)} V^l W_k^{(l+1)} + b_k^{(l+1)}) | k \in (1, \ldots, f)\right)$ (5)

where V^l is the node feature matrix in the lth layer, $\delta(.)$ is the activation function. $W^{(l+1)}$ and $b^{(l+1)}$ are learnable parameters. E_k^l is the kth feature matrix of edges.

Furthermore, we update the edge features in two steps. In the first step, we define a vector $e_{i,j}$ to represent the relationship between nodes i and j:

$e_{i,j}^{(l+1)} = \delta((V_i^{(l+1)} || V_j^{(l+1)} W_{i,j}^{(l+1)} + b_{i,j}^{(l+1)})$ (6)

where $e_{i,j}^{(l+1)}$ is the relation vector between node i and node j of the $(l+1)$th layer, $\delta(.)$ is the activation function, $V_i^{(l)}$ and $V_i^{(l+1)}$ are the node feature vectors of layer l and layer $(l+1)$ respectively, $W_{i,j}^{(l+1)}$ and $b_{i,j}^{(l+1)}$ are learnable parameters. Then, the feature vector $E_{i,j}$ of the edge between the node i and the node j is updated by:

$E_{i,j}^{(l+1)} = Update(E_{i,j}^l, e_{i,j}^{(l+1)})$ (7)

$Update(.)$ represents the update function, $E_{i,j}^l$ represents the edge feature matrix between node i and node j in layer l. Overall, we extend the work of GCN by setting $l = 2$ as the default to form a two-layer EGCN network. To ensure the effectiveness of the model, only the first layer will be set to carry out edge update steps, while the second layer will not carry out edge update steps.

(3) Finish feature extraction

Finally, by this way, we can get the drug's local features (sequence features) C_{drug}, protein's local features (sequence features) $C_{protein}$, drug's global features (molecule structure features) $V_n \in R^{N_n \times f}$ and protein's global features (contact map) $A_a \in R^{r \times f}$. In addition, $A_a \in R^{r \times f}$ and C_{drug} are fed into GLTransformer's decoder. $V_n \in R^{N_n \times f}$ and $C_{protein}$ are fed into GLTransformer's encoder.

(4) GLTransformer module

We construct the GLTransformer structure to extract interactive features, which is shown in Figure 1C and takes drugs and targets as inputs.

(5) GLTransformer's encoder

The GLTransformer's encoder is shown in Figure 1E and takes $C_{protein} \in R^{d_{length} \times \rho}$ and $V_n \in R^{N_n \times f}$ as inputs. The structure of the encoder is constituted by the multi-layer combinations of one-dimensional convolutions and gated linear units. It ensures that effective features can be extracted while preventing the model from overfitting due to complex structures. Then, we can get the feature matrices of drugs and targets after the ith encoder, which is described as follows:

$D_{i+1}^{encoder} = \delta(D_i^{encoder} * W1_i^{drug} + b1_i^{drug})$ (8)

$P_{i+1}^{encoder} = \delta(P_i^{encoder} * W2_i^{prot} + b2_i^{prot})$ (9)

where $\delta(.)$ is the sigmoid function. $*$ is the one-dimensional convolution operation. $W1_i^{drug}, \in R^{k_d \times m \times m}$, $W2_i^{prot} \in R^{k_p \times m \times m}$, $b1_i^{drug}, b2_i^{prot} \in R^m$ are the learnable parameters. k_d and k_p are the filter sizes in encoders, respectively. m is the feature dimension of hidden layers. $D_{i+1}^{encoder} \in R^{P_{length} \times m}$ and $P_{i+1}^{encoder} \in R^{r \times m}$ are the inputs to the ith encoder. In particular, $D_1^{encoder} = C_{protein}$ and $P_1^{encoder} = V_n$. $D_{i+1}^{encoder}$ and $P_{i+1}^{encoder}$ are the outputs of the i th encoder in GLTransformer, where $i \epsilon \{1 \ldots L\}$ and L is the number of encoder layers. Finally, we can get feature matrices of structures for drugs and targets $\left[D_{i+1}^{encoder}, P_{i+1}^{encoder}\right]$ after convolutions by multiple encoders.

(6) GLTransformer's decoder

The GLTransformer's decoder is shown in Figure 1D and takes $C_{drug} \in R^{d_{length} \times \rho}$ and $V_n \in R^{N_n \times f}$ as inputs. A crucial element of the multi-head attention layers is the scale dot attention layers, primarily utilized for extracting the interaction features of proteins and drugs. Attention has three parameters Q, K, and V. their calculation formula are as follows:

$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \times V$ (10)

where Q is the input information of the decoder, and K and V are the output information of the encoder. d_k is the scaling factor. The decoder includes 8 heads of attention, and the formula for calculating the attention of each $head_i$ is as follows:

$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$ (11)

$MutiHead(Q, K, V) = Concat(head_1, \cdots, head_8)W^0$ (12)

Where $b = 8$ and W^O is a parameter that can be learned.

Predictor

In this study, the drug-target prediction is modelled as a binary classification task. All the independent and interactive features learned in previous sections are concatenated and fused to the fully connected dense layers in the downstream:

$\hat{y} = \delta(W_{out}[Features_{Sequence}; Features_{structure}] + b_{out})$ (13)

Where $\delta(.)$ is the sigmoid function, W_{out} and b_{out} are the learnable parameters, and $\hat{y}$ is the predicted label.

For the sets of drug-target pairs and their ground-truth labels in the training datasets, the cross-entropy loss function is defined as the loss function for backpropagation.

$$Loss(\theta) = -\frac{1}{N}\sum_i [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] + \frac{1}{2}||\theta||_2^2 \quad (14)$$

Results and discussion

Evaluation criteria

We adopt six widely used evaluation metrics to evaluate the performance of DCDTI. These metrics include the area under the receiver operating characteristic (ROC) curve (i.e., AUC), area under the precision-recall (PR) curve (i.e., AUPR), accuracy (ACC), precision (PREC), recall (REC), and F1- score (F1).

$$Accuracy(Acc.) = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Precision(Prec.) = \frac{TP}{FP + TP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (18)$$

Where TP, FP, TN and FN represent the number of true positives, false positives, true negatives and false negatives, respectively. The higher the metric value, the better the performance of the model.

Performance of DCDTI

To evaluate the performance of DCDTI for novel DTI prediction under different data distributions, we conduct five-fold cross-validation (5-CV) on DUD-E and Human datasets, respectively. The known drug-target association pairs are regarded as positive samples, whereas negative samples are constructed from the remaining unknown drug-target pairs. Of note, we ensure that an equal number of negative samples is generated as that of positive samples in this experiment. The positive samples and negative samples are constructed as the experimental dataset and taken as input for 5-CV experiments.

In detail, 5-CV randomly divides the data into five subsets and each time one subset is selected as the test set while the remaining parts are used as the training set. To minimize the potential bias of the 5-CV experimental result, the mean and standard deviation values of the metric scores are calculated. The AUPR provides more informative insights, especially for the imbalanced datasets. Thus we use AUC and AUPR as the primary metrics for model performance evaluation.

The results are detailed in Table 2. In particular, the AUC scores of DCDTI on DUD-E and Human are 0.9855 and 0.9714, respectively, whereas the AUPR scores are 0.9104 and 0.9102, respectively. Notably, the AUPR value for DCDTI on the DUD-E dataset (0.9104) exceeds that on the Human dataset. This is primarily due to the higher density of the DUD-E dataset compared to that of the Human dataset.

Table 2. DCDTI main performance on two datasets.

Dataset	Fold	AUC	AUPR
DUD-E	1	0.9854	0.9161
	2	0.9830	0.9226
	3	0.9900	0.8997
	4	0.9773	0.9012
	5	0.9920	0.9123
	Mean	0.9855	0.9104
Human	1	0.9662	0.9023
	2	0.9745	0.9106
	3	0.9712	0.9178
	4	0.9731	0.9120
	5	0.9720	0.9085
	Mean	0.9714	0.9102

Consequently, the model is able to capture more features related to drug-target interactions. AUC in the third fold is slightly higher than that of the other folds. The uneven performance of DCDTI on the third fold demonstrates inferior performance and generalization ability when facing various data with different distributions. Overall,

the satisfactory results indicate that DCDTI can adapt to data with different distributions and validate the availability of DCDTI for potential DTI prediction.

Baseline methods

To evaluate the performance of the proposed model, we compared DCDTI with the five state-of-the-art drug-target prediction models as follows.

(1) TransformerCPI learns the low-dimensional vector representations of drugs and proteins in sequence by using the word2vec algorithm and detects new DTI based on Transformer [15].

(2) GCN-DTI learns the features for each node in a constructed drug-protein pair network by using a GCN and then uses a deep neural network to predict DTI.

(3) Graph Neural Network for Drug-Target Affinity (GraphDTA) models drugs as molecular graphs, uses GCNs to learn drug molecular graph features, and uses CNNs to capture protein feature representations. Finally, the DTI's probability is predicted through the fully connected layer [16].

(4) Deep Convolutional Neural Network for Drug-Target Interaction (DCNNDTI) uses convolutional filters to capture local residue patterns in DTI, uses data as high-level input, constructs model protein features and concatenates drug features. Finally, the DTI's probability is predicted through the fully connected layer [17].

(5) Interactive and Independent Features for Drug-Target Interaction (IIFDTI) employs bidirectional encoder-decoder to extract the interactive features of substructures between drugs and targets, and uses GNNs and CNNs to extract independent features of drugs and targets. Then, all extracted features are fused and input into fully connected dense layers for predicting DTI.

DCNNDTI and TransformerCPI are sequence-based methods (local features), while GCN-DTI and GraphDTA are network-based methods (global features). IIFDTI combines sequence-based and network-based methods. To comprehensively evaluate the DCDTI performance, we run all the compared methods on the DUD-E and Human dataset. We perform comparison through 5-fold cross-validation and obtain the mean AUC and AUPR scores as the main evaluation metric. The corresponding results are shown in Figure 2 and Figure 3. It is worth noting that all baselines have a significant decline on the Human dataset, which indicates that data distribution has a great impact on model performance. The proposed model DCDTI achieves the best performance in all approaches.

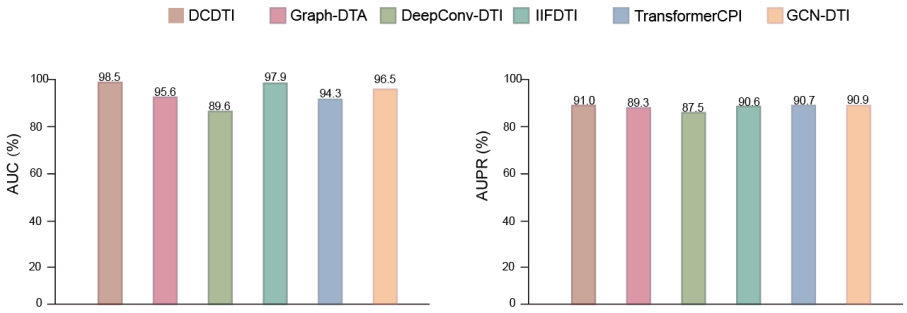

Figure 2. Performance comparison of our method and existing methods on DUD-E dataset.

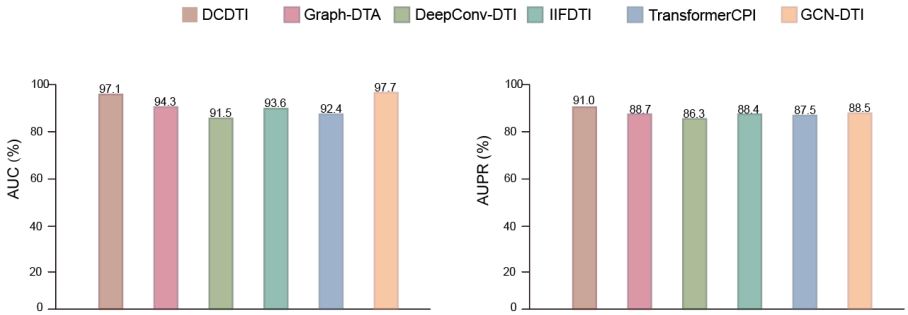

Figure 3. Performance comparison of our method and existing methods on the Human dataset.

The AUC values of DCDTI on DUD-E and Human are 98.55% and 97.14%, respectively. The AUPR values of DCDTI on DUD-E and Human are about 91.00%, respectively. Particularly, for the DUD-E dataset, GCN-DTI obtains the performance second only to DCDTI, and the gaps are 2.00% and 0.10% in terms of AUC and

AUPR, respectively. It is demonstrated that the contribution of the proposed GLTransformer exceeds that of the bidirectional encoder-decoder. GCN-DTI obtains the third-best performance after DCDTI on the DUD-E dataset, suggesting that global structure information captured by GCN is useful for predicting DTI but not as good as combining global structure and local structure information. DCNNDTI and TransformerCPI get the poorest performance across all metrics on DUD-E and the Human dataset, implying the poor performance of only using sequence information and the unavailability in DTI predictions. It is also worth noting that on the Human dataset, DCDTI gets 97.14+% AUC value, while GCN-DTI gets 97.70% AUC value. Despite having a lower AUC than GCN-DTI, the AUPR of DCDTI is higher than that of GCN-DTI.

Hyperparameter analysis

We perform detailed hyperparameter sensitivity experiments on DCDTI with three other hyperparameters because some crucial parameters may affect its performance. We mainly focus specifically on three parameters: the learning rate, dropout and EGCN layer. The corresponding experiment results are shown in Figure 4 and Figure 5.

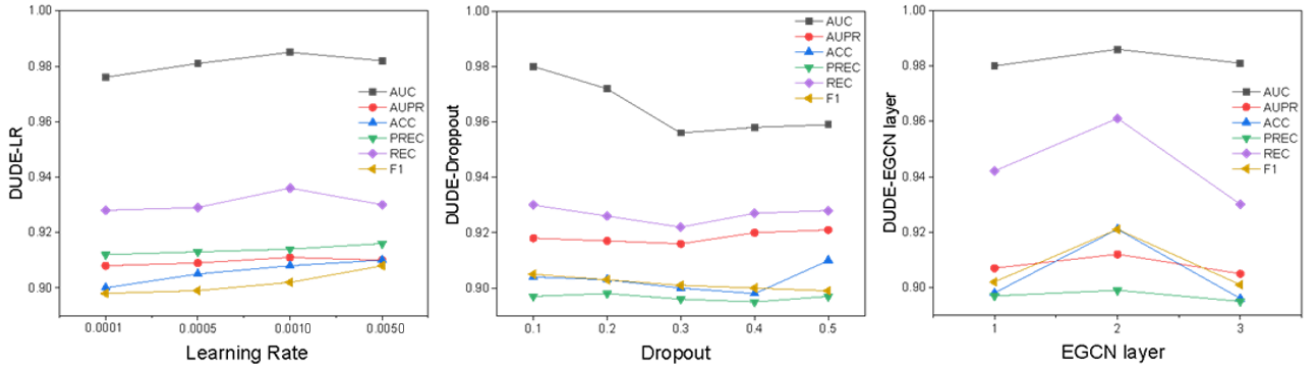

Figure 4. Hyperparameter Sensitivity analysis of DCDTI on DUD-E dataset.

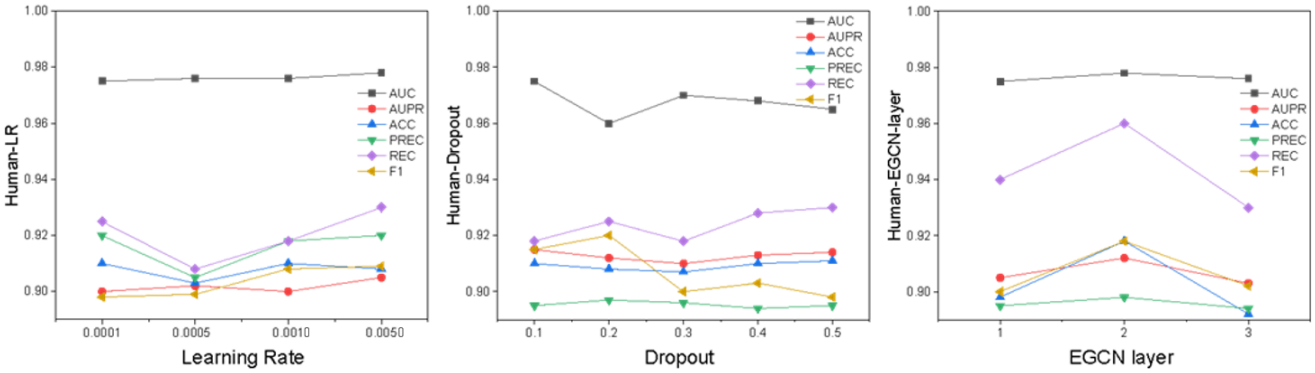

Figure 5. Hyperparameter Sensitivity analysis of DCDTI on the Human dataset.

(1) Effect of the learning rate

The learning rate is a hyperparameter that regulates the extent to which a model is adjusted based on the estimated error. The selection of an appropriate learning rate is challenging as a small value may lead to a long training process, whereas a large value may cause instability in the learning process. Therefore, it is crucial to investigate how different learning rates impact the prediction performance of DCDTI.

From Figure 4, we observe that the performance of DCDTI increases initially and then decreases as the learning rate increases on DUD-E datasets. Specifically, on the DUD-E dataset, the scores are highest. This verifies that a large or a small learning rate will lead to performance degradation. For the Human dataset, the results are stable. We finally set the learning rate to 0.001 for both datasets.

(2) Effect of the dropout

Dropout is a technique for addressing overfitting by randomly dropping units from the neural networks during training. DCDTI chose the values from {0.1, 0.2, 0.3, 0.4, 0.5} and the results are shown in Figure 4 and Figure 5. DCDTI obtains the best scores when the dropout is 0.1 on both datasets.

(3) Effect of the EGCN layers

In addition, a small number of EGCN layers will result in underfitting, whereas a large number can result in overfitting and a long model training time. As a result, DCDTI searches on hidden size from {1, 2, 3}, and we evaluate the model performance under these different EGCN layers. As shown in Figure 4 and Figure 5, DCDTI achieves the best results when the EGCN layer is 2 for datasets DUD-E and Human, respectively.

Ablation study

The proposed DCDTI mainly contains four essential components, i.e., local features (sequence features), global features (molecule features), EGCN, and GLTransformer. Here, the ablation experiments are considered to check the contribution of these key components. The model variants are summarized as follows.

(1) DCDTI-global means DCDTI without global features of drugs and targets. The local features of the drug and target are concatenated and fed into a fully connected neural network.

(2) DCDTI-local denotes DCDTI without local features of drugs and targets. The global features of the drug and target are concatenated and fed into a fully connected neural network.

(3) DCDTI-EGCN means DCDTI without EGCN but with GCN.

(4) DCDTI-GLTransformer denotes the model without GLTransformer. The local features and global features of the drug and target are concatenated and fed into a fully connected neural network.

Table 3 presents the AUC and AUPRC values of the DCDTI and the four variants on the DUD-E dataset. From the table, we can find that the performance of DCDTI is better than that of the other four variants. In particular, DCDTI far outperforms DCDTI-global and DCDTI-local. Compared with the DCDTI-global, DCDTI achieves an improvement of 4.02% and 1.12% in terms of AUC and AUPR, respectively. Compared with DCDTI-local, DCDTI is better than it by 4.95% and 0.79% in terms of AUC and AUPR, respectively. This result clearly indicates that the integration of local and global features helps to improve the prediction performance of DCDTI.

Table 3. Performance of DCDTI and four variants.

Dataset	Method	AUC	AUPR
DUD-E	DCDTI	0.9855	0.9104
	DCDTI-global	0.9453	0.8992
	DCDTI-local	0.9360	0.9025
	DCDTI-EGCN	0.9751	0.9077
	DCDTI-GLTransformer	0.9165	0.8878
Human	DCDTI	0.9714	0.9102
	DCDTI-global	0.9533	0.9006
	DCDTI-local	0.9446	0.8997
	DCDTI-EGCN	0.9645	0.8913
	DCDTI-GLTransformer	0.9153	0.8835

We observe that DCDTI-GLTransformer exhibits significant performance degradation among all variants. Compared with DCDTI, it leads to a drop of 6.90% in AUC and 2.26% in AUPR. This result indicates that the joint aggregation of local and global semantic information via GLTransformer is critical to achieving satisfactory performance, as it enables the model to effectively learn drug-protein interaction patterns. Furthermore, DCDTI outperforms DCDTI-EGCN by 1.04% in AUC and 0.27% in AUPR, respectively.

Case study

To verify the performance of our model generalization and utility, we select Pyridoxal phosphate (DrugBank ID: DB00114) and Aspartate aminotransferase (Uniprot ID: Q2TU84) for prediction on DCDTI. In this section, we use converged models to predict the probability of their interaction with a known drug or target in the dataset.

As shown in Table 3 and Table 4, Table 3 is the result of candidate targets for the new drug Pyridoxal phosphate and Table 4 is the result of candidate drugs for the target

Cytochrome P450 3A4. For predicted outcomes, we rank drug candidates (or targets) according to their predicted scores. Finally, the predicted interactions are verified against the DUD-E database.

We use the drug prediction targets as described. We predict 2,869 targets, and the top 10 targets by prediction scores are shown in Table 4. Three of the predicted proteins have been validated in the Human database. The remaining proteins have also been validated in other literature sources or databases. These results collectively validate our prediction approach. Similarly, the top 10 predicted drug probabilities for targets are shown in Table 5. Three out of the 4,400 candidate drugs we predict have been validated in the database. These three drugs are ranked within the top 10. The results demonstrate that the DCDTI can screen out interaction candidates from a large number of drug and target pairs, which is of great significance for drug screening.

Table 4. The predicted candidate targets for the new drug Pyridoxal phosphate.

Rank	Target name	Evidence
1	Alanine - glyoxylate aminotransferase	Human
2	Cystathionine beta-synthase	DUD-E
3	Ornithine decarboxylase	PubMed: 18816584
4	Histidine decarboxylase	Human
5	Calstabin-1	Human
6	Cholesterol 25-hydroxylase	PubMed: 19818325
7	Phosphoserine aminotransferase	Unknown
8	Cytochrome P450 1A2	PubMed: 18814214
9	Phosphorylase	Unknown
10	CYPIID6	PubMed: 12459873

Table 5. The predicted candidate drugs for target Cytochrome P450 3A4.

Rank	Drug name	Evidence
1	Pradefovir mesylate	Human
2	Flunisolide	Human
3	Cevimeline	DUD-E
4	Bortezomib	PubMed: 240513
5	Fluconazole	PubMed: 24121043
6	Dofetilide	Unknown
7	Pantoprazole	DUD-E
8	Eletriptan	Unknown
9	Nelfinavir	DUD-E
10	Indinavir	PubMed: 26232224

Model visualization

To further explore the learning capabilities of DCDTI, we perform visualization experiments on two datasets. In particular, if there is an association relationship between one drug and one target, this drug-protein pair will be labelled with a positive pair; otherwise, it will be labelled with a negative pair. All the embeddings of drug-target pairs are then projected into a two-dimensional space using Uniform Manifold Approximation and Projection (UMAP), respectively.

As shown in Figure 7 and Figure 8, the predicted DTI by DCDTI are clearly clustered into different groups, i.e., positive and negative links are distinguished well, which suggests that DCDTI thus consistently and reliably allows accurate representations for DTI prediction. It should be noted that there is still a mixture of colored and red points in certain areas, indicating the presence of an extremely challenging decision boundary in the DTI

prediction task. This observation further confirms the discriminative nature of the learned embeddings of drug-target pairs, thereby improving the credibility of DCDTI in predicting DTI.

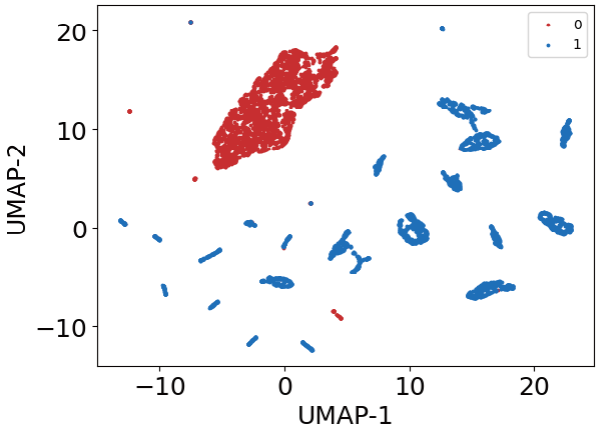

Figure 6. Visualization of UMAP on DUD-E dataset.

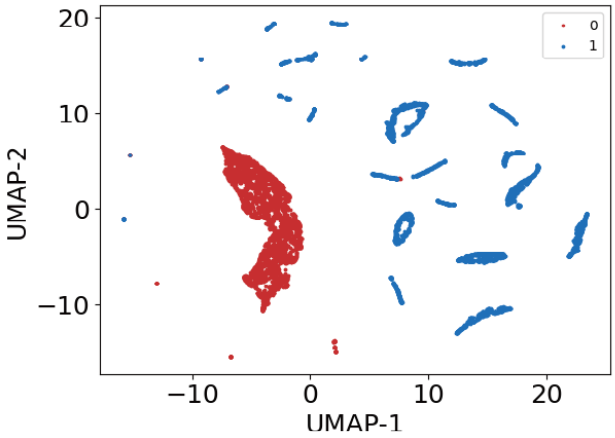

Figure 7. Visualization of UMAP on Human dataset.

Discussion

In this study, by combining local attributes with global attributes for DTI prediction, our model integrates the advantages of sequence-based and graph-based methods. That is, it can not only adaptively obtain drug (protein) characterization, but also fully utilize the structural information of drugs and proteins. FCS text mining is used to acquire local features. Despite full consideration of the various attribute characteristics of drugs and targets, the local feature extraction method is still slightly insufficient. In terms of the FCS text mining algorithm, we mainly consider the bag-of-words method. However, this method treats the text as an unordered set of words and ignores the order information and semantic similarity of words. This results in the fact that the model cannot capture the contextual relationship and grammatical structure between words and loses important information in the text. Further studies are needed to obtain optimal text mining algorithms and local feature learning methods.

Conclusion

The identification of active candidate compounds for target proteins can facilitate the development of drugs in molecular receptor targeting and complex disease therapy. In this work, we demonstrate that DCDTI can efficiently improve the accuracy and utility of DTI identification based on the deep neural network. DCDTI is a Dual-Channel Neural Network for Drug-Target Interaction prediction model based on GLTransformer. The DCDTI method contains three main modules: the multi-modal data input module, the feature extraction module and the GLTransformer module. We utilize AUC, AUPR, ACC, and F1-score metrics to evaluate the performance of DCDTI. According to the experimental results, our model achieves the highest AUC and AUPR values of 0.9855 and 0.9104. Moreover, the ablation study is conducted to explore why and how DCDTI can outperform other state-of-the-art DTI methods. The limitations and future work are also discussed. We believe that our proposed model could pave the way for DTI prediction and foster the development of drug-related studies.

Funding

This work was financially supported in part by the Natural Science Basic Research Project of Shaanxi Province (Grant No. 2025JC-YBMS-773).

Authors' contributions

Yongbin Zeng conceived the basic idea. Ping Zhang designed, implemented the model and performed methods comparison. Ping Zhang wrote the the manuscript. Yongbin Zeng revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Bleakley, K., Yamanishi, Y. (2009) Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*, 25(18), 2397-2403.

[2] Cheng, F., Zhou, Y., Li, J., Li, W., Liu, G., Tang, Y. (2012) Prediction of chemical-protein interactions: Multitarget-QSAR versus computational chemogenomic methods. *Molecular BioSystems*, 8(9), 2373-2384.

[3] Mohamed, S. K., Nováček, V., Nounu, A. (2020) Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, 36(2), 603-610.

[4] Olayan, R. S., Ashoor, H., Bajic, V. B. (2018) DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics*, 34(7), 1164-1173.

[5] Wang, Y., Zeng, J. (2013) Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics*, 29(13), i126-i134.

[6] Yuan, Q., Gao, J., Wu, D., Zhang, S., Mamitsuka, H., Zhu, S. (2016) DrugE-Rank: Improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*, 32(12), i18-i27.

[7] Zong, N., 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.

[8] Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., Li, Z., Luo, X., Chen, K., Jiang, H., Zheng, M. (2020) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 63(16), 8749-8760.

[9] Zhang, Z., Guan, J., Zhou, S. (2021) FraGAT: a fragment-oriented multi-scale graph attention model for molecular property prediction. *Bioinformatics*, 37(18), 2981-2987.

[10] Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., Barzilay, R. (2019) Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8), 3370-3388.

[11] Xuan, P., Zhang, Y., Cui, H., Zhang, T., Guo, M., Nakaguchi, T. (2021) Integrating multi-scale neighbouring topologies and cross-modal similarities for drug-protein interaction prediction. *Briefings in Bioinformatics*, 22(5), bbab119.

[12] Zhao, T., Hu, Y., Valsdottir, L. R., Zang, T., Peng, J. (2021) Identifying drug-target interactions based on graph convolutional network and deep neural network. *Briefings in Bioinformatics*, 22(2), 2141-2150.

[13] Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., Luo, X., Chen, K., Jiang, H., Zheng, M. (2020) TransformerCPI: Improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16), 4406-4414.

[14] Öztürk, H., Özgür, A., Ozkirimli, E. (2018) DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics*, 34(17), i821-i829.

[15] Liu, H., Sun, J., Guan, J., Zheng, J., Zhou, S. (2015) Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics,* 31(12), i221-i229.

[16] 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.

[17] Cheng, Z., Zhao, Q., Li, Y., Wang, J. (2022) IIFDTI: Predicting drug-target interactions through interactive and independent features based on attention mechanism. *Bioinformatics*, 38(17), 4153-4161.