IJE TRANSACTIONS B: Applications Vol. 38 No. 08, (August 2025) 1737-1748

International Journal of Engineering

Journal Homepage: www.ije.ir

Predicting Drug-target Affinity by Discovering Pairwise Interactions Using Cross Attention Network

S. Falahati[a], F. Zamani*[a], M. A. Khodamoradi[b]

[a] *Babol Nooshirvani University of Technology, Electrical and Computer Engineering Department, Babol, Iran*
[b] *Universidade NOVA de Lisboa, NOVA School of Science and Technology (FCT NOVA) / Uninova, Center of Technology and Systems, Portugal*

PAPER INFO

Paper history:
Received 24 September 2024
Received in revised form 16 November 2024
Accepted 18 November 2024

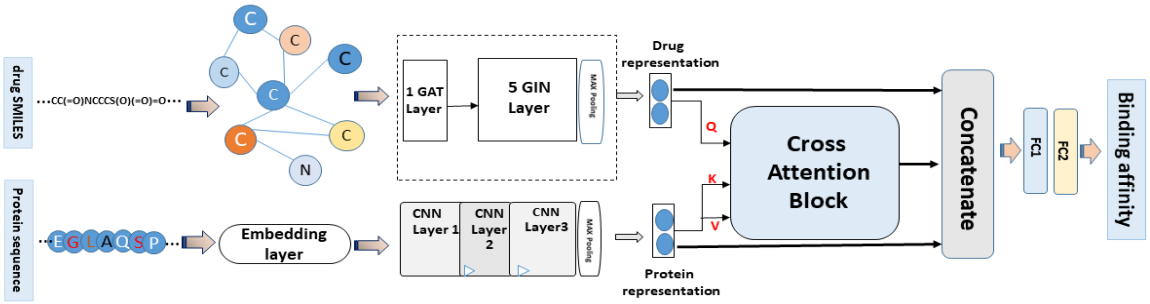
Keywords:
Drug Discovery
Cross-Attention Network
Drug-target Interaction
Binding Affinity Prediction
Graph Isomorphism Networks
Graph Attention Networks

ABSTRACT

In drug discovery, Drug-Target Affinity (DTA) is considered as a vital step, as it helps identify the most promising drug candidates in the development process. Since the structure and function of drug and target molecules must be considered, as well as their complex and nonlinear interactions, DTA prediction is a challenging task. The aim of this study is to propose a novel DTA prediction framework that leverages the strengths of Cross-Attention Networks (CANs) using Graph Neural Networks (GNNs). However, representing graphs using GNNs keeps their 3D structural information. They are not fully exploited by existing attention-based approaches. Our framework uses CAN to capture a more accurate representation of the drug-target pair by analyzing how different parts of a drug molecule interact with specific regions of the protein. We used GIN and GAT in a sequential architecture to capture both local and global structural information of drug graph molecules. We evaluate the performance of the proposed method on two benchmark datasets, Davis and KIBA. The performance is promising while it outperforms many state-of-the-art methods in terms of mean square error (MSE) and concordance index (CI). Specifically, for the Davis dataset, we achieve MSE of 0.222 and CI of 0.901, while for KIBA, we obtained MSE of 0.144 and CI of 0.883. Our method increases the interpretability and specificity of interaction analysis, providing deeper insight into the drug discovery process and providing valuable explanations for the predicted DTA. The code of our study is available at: https://github.com/fsonya88/CAN-DTA.

doi: *10.5829/ije.2025.38.08b.01*

Graphical Abstract



1. INTRODUCTION

One of the main goals of pharmaceutical science is to prepare drugs that can help treat various human diseases. However, the way to reach this goal is very long and expensive. To produce a new drug, it is necessary to go through various steps, including drug target identification, molecular design, laboratory and clinical trials. The process usually takes between 10 and 15 years and costs between $0.8 and $1.5 billion (1). Drug

*Corresponding Author Email: zamani@nit.ac.ir (F. Zamani)

Please cite this article as: Falahati S, Zamani F, Khodamoradi MA. Predicting Drug-target Affinity by Discovering Pairwise Interactions Using Cross Attention Network. International Journal of Engineering, Transactions B: Applications. 2025;38(08):1737-48.

repurposing is an approach to accelerate the drug discovery process by seeking to discover new uses for an existing drug that is not currently prescribed (2). DTA describes the binding strength between drug molecules and protein targets. It helps the drug-repurposing process by predicting unknown relationships between drug molecules and the protein target of the disease for which the drug was repurposed (3). Drug molecules and protein target binding prediction is challenging as it involves complex molecular interactions and large-scale data analysis. With recent advances in the field of artificial intelligence (AI), it is possible to use this technology in drug discovery (4). AI can play an important role in improving the efficiency and reducing the time and costs of pharmaceutical research. Some of the benefits of using AI include reducing human error in drug design, increasing speed and accuracy in drug design, increasing drug effectiveness, predicting designed drug risks and side effects, and identifying new drug molecules (5). AI uses various methods to predict the interaction between drugs and proteins. These methods include convolutional neural network (CNN) (6) graph neural network(GNN) (7, 8) support vector machine (SVM), and random forest (RF) (9).

Different approaches for DTA prediction can broadly be categorized into three types: sequence-based, feature-based, and graph-based methods (10-13). Sequence-based methods use the sequences of the target protein and the drug as inputs, and encode them into numerical vectors using various techniques, such as one-hot encoding, n-gram encoding, or word embedding. Then, they use machine learning or deep learning models, such as SVM, RF, or CNN, to learn the features and patterns of the sequences and predict the binding affinity (14). An example of a sequence-based model for DTA prediction is DeepDTA (15), which used CNN to learn representations from the protein sequences and Simplified Molecular Input Line Entry System (SMILES) (16) strings of the drugs. However, sequence-based methods may not capture the structural and chemical properties of the molecules, which are important for DTA prediction. Feature-based methods extract features from drugs and targets, such as fingerprints or descriptors, and use them as inputs for machine learning models. Alternatively, they use the three-dimensional structures of the drug and the target protein as inputs and simulate the docking or binding process between them using various techniques. However, feature-based methods may not capture the dynamic and context-dependent nature of the interactions, which may vary depending on the environment and the conformation of the molecules. Graph-based methods use graph representations of drugs and proteins, which can capture their structural and chemical properties more effectively than strings, matrices, or 3D structures (17-19). These methods use GNNs to learn from the graph-structured data, which can handle both labeled and unlabeled data, using supervised or unsupervised learning methods. GNNs can use different kinds of graphs, such as directed, undirected, weighted, or attributed graphs, depending on the data and the task. An example of a graph-based model for DTA prediction is Graph DTA (20), which uses different types of GNNs to represent the drugs as molecular graphs and CNN to represent the protein vector.

This paper addresses the following research question: How can we use deep learning to predict drug-target binding affinity more accurately and efficiently. Our hypothesis is that by combining GATs (21), GINs (22), and CANs (23, 24), we can capture both the local and global structure of the drug and the pairwise interactions between the drug and the protein, which can enhance the representation of the drug-target pair and improve the DTA prediction performance. GATs are graph neural networks that use attention mechanisms to assign different importance to different neighboring nodes when aggregating their features. GINs are graph neural networks that can learn expressive and invariant features from graph-structured data. CANs use cross-attention to compute the attention weights between the drug and the target embeddings, which allows the model to focus on the most relevant parts of the sequences for the binding affinity prediction. By using these three approaches together, our proposed model can capture both the local and global features of the graph structure and node attributes, as well as the complex and nonlinear interactions between drug and target molecules. This study aims to propose a method for DTA prediction based on GNNs and Cross-Attention Networks to enhance performance and accuracy. To achieve this aim:

- We propose a novel deep learning-based DTA model that incorporates the deep representations of drugs and targets by GAT, GIN, and CNN. We used a combination of GATConv and GINConv to represent drugs as molecular graphs, which can capture both the local and global features of the graph structure and node attributes more effectively than traditional graph convolutional networks.
- Deep graph-based representations of drugs and targets are not fully exploited by existing attention-based methods. To gain a deeper understanding of how the drug and target protein interact, our model incorporates a single-head cross-attention block. By employing CAN, our framework analyzes how different parts of a drug molecule interact with specific regions of the protein, leading to a more accurate understanding of the drug-target pair and its binding affinity.
- We combine the drug, target, and cross-attention embedding into a single vector and pass it through several fully connected layers to produce the final

output. We train and evaluate our model on benchmark datasets and show competitive performance with state-of-the-art methods.
- Our model uses protein sequences rather than complex protein structures, significantly increasing computational efficiency without sacrificing performance. This streamlined approach is especially valuable for analyzing large drug-target interaction datasets.

2. RELATED WORKS

Many machine learning-based and deep learning-based methods have been developed for DTA predictions. KronRLS is a method that uses the similarity-based approach and the Kronecker product kernel to capture the joint similarity between drugs and targets (25). In this model Smith-Waterman algorithm is used for aligning sequences. It then uses a regularized least-squares algorithm to learn a regression model that predicts drug-target binding affinity. SimBoost is a machine-learning-based model that predicts the continuous binding affinities of drugs and targets using gradient-boosting machines (26). SimBoost used Smith-Waterman algorithm for aligning sequences and PubChem similarity for drugs then constructs features for each drug, target and drug–target pair based on their properties and similarities. Deep-learning methods employ various types of neural networks to learn the patterns and interactions between drugs and targets from their sequences. For instance, DeepDTA is a deep learning method that utilizes CNNs to encode drug SMILES strings and protein amino acid sequences. It then uses a fully connected layer to estimate drug-target binding affinity. WideDTA is another deep-learning method that employs word embedding to represent drug SMILES strings and protein sequences (27). It then utilizes CNNs to encode the drug and target embedding. It then utilizes CNNs to encode the drug and target embedding. Attention DTA predicts the binding affinities of drugs and targets using sequence-based deep learning with an attention mechanism (28). It uses an attention mechanism to focus on the important subsequences in the drug SMILES strings and protein sequences that are relevant for binding affinity prediction.

GANsDTA is a semi-supervised based model which uses two GANs to extract features from protein and drug sequences (29).

DeepCDA combined CNN and LSTM to capture the local and global features of drug and protein sequences and used a two-sided attention mechanism to encode the interaction strength of the substructures (30). DeepCDA also applies an adversarial domain adaptation technique to learn a feature encoder network in the test domain. DeepMHADTA predicts drug target binding affinity based on a multi-head attention using protein sequences and structures as well as drug sequences and their structural fingerprints (31). It treats drug fingerprints and protein structures as sequences, using CNN for feature extraction and multi-head attention to increase prediction accuracy.

DeepGS predicts the binding affinity of drugs and targets using deep neural networks (32). DeepGS extracts the local chemical context from the drug and target sequences using advanced embedding techniques and models the molecular structure of drugs using graph convolutional layers. DeepGS then combines the drug and target embeddings using a bilinear layer and predicts the binding affinity using a fully connected layer. The study shows that DeepGS outperforms state-of-the-art methods on four benchmark datasets and provides interpretable results by highlighting the important regions of the drugs and targets that contribute to binding affinity. DeepGLSTM predicts the binding affinity between drugs and targets using a deep graph convolutional network and LSTM (33). DeepGLSTM encodes the drug and target sequences using LSTM and models the molecular structure of the drugs using graph convolutional layers. DeepGLSTM then combines the drug and target embeddings using a bilinear layer and predicts the binding affinity using a fully connected layer.

MGraphDTA uses a deep multi-scale graph neural network for drug representation and a sequence based multi-scale convolutional neural network for protein representation, creating a combined representation that enables detailed insights into drug-target interactions (34). KC-DTA uses the k-mer method to extract features from protein data while drugs are modeled by graph (35). GDilatedDTA introduces a graph dilation strategy in convolutional layers to enhance the feature representation of drug-target pairs (36). DGraphDTA enhances drug-target affinity prediction by constructing separate graphs for drug molecules and proteins, using predicted contact maps from protein sequences to better capture structural features (37).

HiSIF-DTA utilizes a hierarchical framework to fuse multi-level semantic information from protein graphs (38). This approach integrates both basic structural information (such as contact maps) and more complex functional information (like protein-protein interaction networks) into a layered protein graph. WGNN-DTA constructed weighted protein graphs biased on contact map to better represent protein structure (39).

TransVAE-DTA integrates variational autoencoders and transformers for protein representation as well as transformers for drug representation, capturing long-distance dependencies to improve DTA prediction (40).

GTAMP-DTA combines graph transformer with an attention mechanism to focus on the most important relevant features (41). It predicts binding affinities using

the extracted features in combination with pre-trained embedding of proteins and drugs.

Attention MGT-DTA combines graph transformer and attention mechanism for drug and protein representation (42). It uses pre-trained protein sequence embedding as well as 3D protein graphs which are integrated using a cross attention module.

The field of DTA prediction has advanced significantly with the application of machine learning and deep learning models that leverage various techniques, such as attention mechanisms and graph Neural networks. These techniques enable the models to capture the complex and diverse features of drug-target interactions and accurately estimate their binding affinities. Consequently, these models facilitate the process of drug discovery and development by providing novel insights and candidates for potential therapies. A review of all related works is provided in Table 1.

3. MATERIALS AND METHOD

3. 1. Datasets The proposed model was evaluated on two widely used datasets for DTA prediction: DAVIS (43) and KIBA (44). These datasets contain binding affinity scores for drug-target pairs, measured by different metrics. DAVIS measures the binding affinity by computing the dissociation constant K_d, which is derived from Equation 1 (43).

$$PK_d = -\log \frac{K_d}{10^9} \quad (1)$$

KIBA measures the binding affinity by the KIBA score, which is derived from various sources, such as inhibition constant K_i, dissociation constant K_d, or half-maximal inhibitory concentration IC50. Both datasets represent the drugs by SMILES strings and the targets by amino acid sequences. The DAVIS dataset has 68 drugs, 442 targets, and 30,056 interactions. The KIBA dataset has 2,111 drugs, 229 targets, and 118,254 interactions. Table 2 presents the detailed information of these two datasets. The datasets split into training and testing sets, with 80% for training and 20% for testing, following the same data preprocessing steps as previous studies (15, 20).

3. 2. Model Architecture The proposed model predicts the binding affinity of DTA using 1D sequential and molecular information as input, such as protein

TABLE 1. A summary of the related works on DTA prediction is presented in this table. The method they used to represent drugs and targets, which can be sequential or graphical, is listed in columns two and three

Model	Drug Representation	Protein Representation
KronRLS (25)	Sequence(Smith Waterman)	Sequence(Smith Waterman)
SimBoost (26)	Sequence(Smith Waterman)	Sequence(Pubchem-Sim)
DeepDTA (15)	Sequence(CNN)	Sequence(CNN)
WideDTA (27)	Sequence(CNN)	Sequence(CNN)
AttentionDTA (28)	Sequence(CNN+Attention)	Sequence(CNN+Attention)
GANsDTA (29)	Sequence(GAN)	Sequence(GAN)
DeepCDA (30)	Sequence(CNN+LSTM)	Sequence(CNN+LSTM)
DeepMHADTA (31)	Sequence(Attention)	Sequence(CNN)
DeepGS (32)	Graph (GAT+BiGRU)	Sequence(CNN)
GraphDTA (20)	Graph(GNN)	Sequence(CNN)
DeepGLSTM (33)	Graph(GNN)	Sequence(BiLSTM)
MGraphDTA (34)	Graph(GNN)	Sequence(CNN)
KC-DTA (35)	Graph(GNN)	Sequence(CNN)
GDilatedDTA (36)	Graph(GNN)	Sequence(BiLSTM)
DGraphDTA (37)	Graph(GNN)	Graph(GNN)
HiSIF-DTA (38)	Graph(GNN)	Graph(GNN)
WGNN-DTA (39)	Graph(GNN)	Graph(GNN)
TransVAE-DTA (40)	Sequence(Transformer)	Sequence(VAE+Transformer)
GTAMP-DTA (41)	Graph(Transformer with Attention)	Graph(Transformer with Attention)
AttentionMGT-DTA (42)	Graph(Transformer with Attention)	3D-Graph(Transformer with Attention)+Sequence(Pre-trained embedding)

TABLE 2. Benchmark datasets

Data set	Drug	Protein	Interaction	Train samples	Test Samples
DAVIS	68	442	30056	25046	5010
KIBA	2111	229	118254	98545	19709

sequences and SMILES strings. The model consists of three main components: a drug graph encoder, a protein sequence encoder, and a cross-attention network. The drug graph encoder employs a combination of GATConv and GINConv to represent the drugs as molecular graphs. The protein sequence encoder uses an embedding layer followed by CNN layers to represent the protein as embedding vectors. Finally, the cross-attention network, which is based on a single-head attention block, computes the attention weights between the drug representation vector and the protein representation vector. The model combines the drug, target, and cross-attention embeddings and passes them through several fully connected layers to get the final output. Figure 1 illustrates the architecture of proposed medel, it consists of three main components: a drug graph encoder, a protein sequence encoder, and a cross-attention network.

3. 2. 1. Drug Graph Encoder The medical SMILES string is provided as an input to the drug graph encoder. SMILES is a notation that encodes the molecular structure of a drug as a string of characters. The encoder converts the SMILES strings into molecular graphs using the RDKit (45), which is an open-source cheminformatics software. In the molecular graphs, each node represents an atom and each edge represents a bond between two atoms. To represent any chemical molecule as a graph, we need to assign attributes to each node (atom) in the graph. These features must capture certain chemical properties of the atom relevant to target chemical interactions.

In this paper, we followed the approach of GraphDTA, the node features include the atom symbol, which indicates the type of the atom, such as carbon, oxygen, nitrogen, etc. the number of adjacent atoms, the number of adjacent hydrogens, which indicates how many hydrogen atoms are attached to the node, the implicit valence, which indicates the number of electrons that the node can share with other atoms and aromaticity determines whether the node is part of an aromatic ring.

The drug graph encoder then applies a GATConv layer followed by five GINConv layers to update the node features by aggregating the features of their neighbors.

GATConv is a graph convolution layer that employs the attention mechanism to assign different weights to the node and its neighbors. The GATConv layer inserts a set of nodes into the graph and applies a linear transformation through a weighted matrix A to each node. For each node i in the graph, GATConv computes the interaction of node i and its neighbors and the attention coefficients between node i and node j as follows (21):

$$\alpha(Ax_i\ , Ax_j) \tag{2}$$

This is the weight of node j contribution to node i representation. These weights are scaled by a soft-max function and then used to calculate the new features for nodes as follows (21):

$$\sigma\left(\sum_{j\,\epsilon\,Ni}\ \ \alpha_{ij}\ Ax_j\ \right) \tag{3}$$

This means that σ is non-linear activation, and α_{ij} are the coefficients that adjust the importance of each pair of nodes.

GIN is a graph neural network model that can learn the characteristics of nodes and their neighbors as a powerful representation of an aggregate graph. GIN uses a multi-layer perceptron (MLP) and a sum aggregator to iteratively update the attributes of nodes. A sum aggregator is a simple but effective way to combine the features of the node and its neighbors, by adding them together with some weights, the sum aggregator can capture information about multiple groups and is injective, meaning that it can distinguish different node neighborhoods. In addition, GIN introduces a learnable

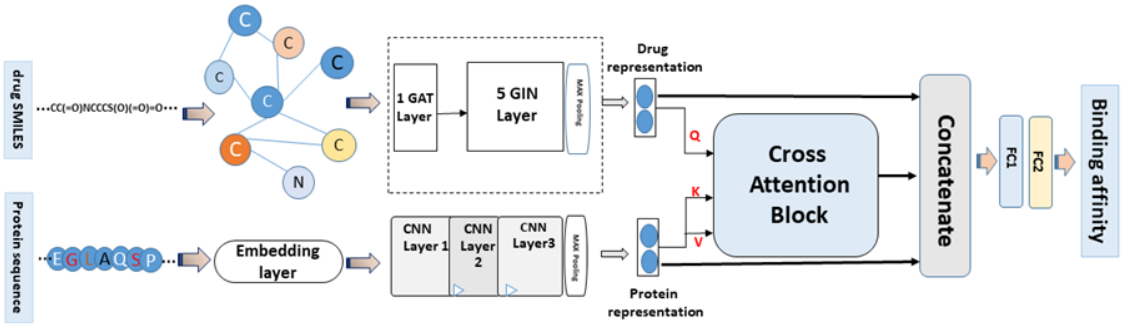


Figure 1. The proposed model

parameter ε that controls how a node's own feature contributes to the set. This parameter is important for the expressiveness of GIN, as it allows the model to approximate any function that is invariant to the permutation of the nodes. The process of GIN can express as follows (22):

$$MLP((1+\varepsilon)x_i + \sum_{j\epsilon Ni} x_i) \quad (4)$$

where Ni are the neighbors of node i, The node attribute of the i-th node is x_i, and it is used to update the node features by aggregating the features of the neighboring nodes in each layer. There is also a batch normalization layer with the ReLU activation function after each MLP layer. GIN can repeat this process for several layers so that each node can capture the information from a larger neighborhood. Finally, GIN can obtain a representation of the whole graph by applying another function, such as a sum or a max, to the feature vectors of all the nodes. This graph representation can serve various tasks; such as graph classification or regression. The drug graph encoder then obtains a fixed-length vector representation of the drug graph through a max-pooling layer.

3. 2. 2. Protein Sequence Encoder Each protein target was represented as a sequence of amino acids. One-hot encoding (46) is used to convert each amino acid to a binary vector. Each sequence is also padded or truncated to a fixed length of 1000 amino acids. Tree CNN layers are used to encode the protein sequence. CNN is a type of neural network that uses convolutional filters to extract local features from the input. CNN consists of multiple convolutional layers, each of which applies a set of filters with a certain kernel size to the input, and then applies a non-linear activation function and a pooling operation to reduce the dimensionality. A dropout layer (47) is also applied to prevent overfitting. After applying CNN, a 32-dimensional feature vector was obtained for the protein sequence. The output is flattened and passed through a linear layer with a ReLU activation (48) and a dropout layer to obtain a 128-dimensional feature vector for the protein target.

3. 2. 3. Cross-Attention Network In addition to compute embeddings of drug and target efficientely, it is equally important for drugs and proteins to engage in mutual attention and information exchange. This interaction is vital, especially considering that substructural elements predominantly influence drug-target affinities. In these cases, the complementarity between specific regions of the drug and protein is a key factor in ensuring a successful binding process, which is central to their biological interaction.

To effectively capture the interactions the drug molecule and the protein target, a cross-attention mechanism was applied to the embedded drug and target vectors. Cross attention is a type of attention mechanism that computes the attention weights between two types of input, and uses them to obtain a weighted combination of the features of one input based on another. Cross-attention consists of three components: query, key, and value. The query is the input that is attended to, the key is the input that is used to compute the attention weights, and the value is the input that is used to aggregate the features. The output of the drug graph encoder is the query, and the output of the protein sequence encoder is both the key and the value. The attention block then computes the attention scores as follows (49):

$$Attention\ score(Q,K,V) = Softmax(\frac{QK^t}{\sqrt{C/_D}}) \quad (5)$$

The attention scores are then multiplied with the value vectors as Equation 6 to obtain a weighted vector representation of the drug and target embeddings (23).

$$Cross\ Attention(Q,K,V) = Attention\ score\ \times V \quad (6)$$

By using a single-head attention block, the model can learn the local and global substructure interaction information of the drug-target pair. Finally, a concatenation of the drug embeddings, target embeddings, and cross-attention scores are passed through a dropout layer to prevent overfitting.

3. 2. 4. Output Block The output block consists of two fully connected layers with ReLU activation functions and dropout layers in between. The first layer takes the concatenated drug-target feature vector as input and outputs a vector of size 1024. The second layer further reduces the dimensionality to 256. The final layer is a linear layer that maps the vector to a scalar value, which represents the predicted binding affinity score. The linear activation function is suitable for the regression task, as the binding affinity can take any real value. The dropout layers are used to prevent overfitting by randomly dropping out some units during training. The output of the model is a score that reflects the strength of the interaction between the drug molecule and the protein target.

4. RESULTS AND DISCUSSION

4. 1. Evaluation Metrics We formulate the prediction of drug-target binding affinity in this paper as a regression problem, Concordance Index (CI) (50) and Mean Squared Error (MSE) (20) are two ways of measuring the accuracy of a regression model, which is a type of model that predicts a continuous value, such as drug-target binding affinity. CI and MSE have different properties and advantages, and they can be used together to evaluate the performance of a regression model.

In this paper, CI and MSE are used as evaluation metrics. CI measures how well the predicted values rank the true values, while MSE measures how close the predicted values are to the true values. A higher CI and a lower MSE indicate better performance. CI and MSE can be calculated using the following equations (50):

$$CI = (1/z)\, \Sigma\, h\,(b_{xi} - b_{yi}) \quad (7)$$

where z is the total number of pairs, h is an indicator function that returns 1 if the pair is concordant, -1 if the pair is discordant, and 0 if the pair is tied, and b_{xi} and b_{yi} are the predicted and true values of each pair, respectively. The step function is shown in Equations 8, 9 (20).

$$h(x) = \begin{cases} 0 \ if \ x < 0 \\ 0.5 \ if \ x = 0 \\ 1 \ if \ x > 0 \end{cases} \quad (8)$$

$$MSE = (1/n) * \sum_{i=1}^{n} (x_i - y_i)^2 \quad (9)$$

where n is the number of observations, x_i is the predicted value of each observation, and y_i is the true value of each observation.CI and MSE are complementary metrics that can be used together to evaluate the performance of a regression model. CI can measure the rank correlation between the predicted and true values, while MSE can measure the absolute error between the predicted and true values. A good regression model should have a high CI and a low MSE, which means that the predictions are both consistent and accurate with the true values.

4. 2. Hyper Parameters and Experimental Setup

The proposed model is implemented with PyTorch (51), which is a deep-learning framework that provides a flexible and efficient way to build and train neural networks. The model follows the same hyper-parameter settings as GraphDTA (20) for fair comparison; Table 2 summarized the hyper-parameters of the experiments, such as learning rate, batch size, and number of GNN layers. The filter sizes of the three convolutional layers in the protein sequence encoder are set to 32, 64, and 96, respectively; in oredr to capture different levels of features from the amino acid sequences. The data is split into 80% for training and 20% for testing. The model trained on 80% of the training data and the best model was selected based on the lowest mean squared error on the validation data, which is 20% of the training data.

4. 3. Comparison with Existing Methods

In this section, we reported the results of our proposed model for DTA prediction on the Davis and KIBA datasets, which are benchmarks. We use MSE and CI as the evaluation metrics.

Performance comparisons between the proposed model and existing models are provided in Table 4 (Davis dataset) and Table 5 (KIBA dataset).

TABLE 3. Hyper-parameters of our experiments

Hyper Parameter	Value
Number of Epochs	1000
Batch size	512
Learning rate	0.0004
Optimizer	Adam
Number of GAT Layers	1
Number of GIN Layers	5
Number of head	1
Number of Fully connected Layers	2

We compare our model with several baselines, including KronRLS (25), SimBoost (26), DeepDTA (15), WideDTA (27), AttentionDTA (28), GANsDTA (29), DeepCDA (30), DeepMHADTA (31), DeepGS (32), GraphDTA (20), DeepGLSTM (30), KC-DTA (35), MGraphDTA (34), and GDilatedDTA (36). Our model achieved the lowest MSE and the highest CI among most of these methods, indicating that it can accurately predict the DTA scores. Our model outperforms GraphDTA, which also uses GINs to encode the drug graphs, by 0.007 in MSE and 0.008 in CI, demonstrating the effectiveness of using GATs to capture the local structure of the drug graphs. Our model also outperforms DeepGLSTM, which uses CNNs and LSTMs to encode the drug and protein sequences, by 0.010 in MSE and 0.006 in CI, showing the advantage of using graph representations over sequence representations. Moreover, our model surpasses by 0.022 in MSE and 0.006 in CI, indicating that our cross-attention mechanism can better capture the pairwise interactions.

As mentioned in the related work section, WGNNDTA, DGraphDTA, HiSIF-DTA, and Attention MGT-DTA employ more complex graph representations of proteins, whereas we used protein sequences.

Similarly, GTAMP-DTA and TransVAE-DTA utilize complex transformer models, which we did not incorporate.

Additionally, MGraphDTA is a deep Graph Neural Network with 27 layers, while our model is a very shallower one.

Despite these differences, the results of the proposed model, which relies on protein sequences rather than complex graph-based representations, are not significantly different from the reported results and are even superior in some cases, in MSE or CI while our model offers better runtime performance. Our model offers a substantial advantage in terms of computational efficiency, making it a practical choice for large-scale drug discovery applications.

Our paper aims to demonstrate the effectiveness of cross-attention. It is crucial for drugs and proteins to

engage in mutual attention and exchange information. This interaction is crucial, especially since substructural elements largely influence drug-target affinities. In such cases, the complementarity between specific regions of the drug and protein is crucial for successful binding, which plays a fundamental role in their biological interaction.

TABLE 4. Performance of our model and the baselines on the Davis dataset

MODELS	Drug	Protein	MSE	CI
KronRLS (25)	Smith-Waterman	PubChem-Sim	0.379	0.871
SimBoost (26)	Smith–Waterman	PubChem-Sim	0.282	0.872
DeepDTA (15)	CNN	CNN	0.261	0.878
GANsDTA (29)	GAN	GAN	0.276	0.881
DeepGS (32)	GAT+BiGRU	CNN	0.252	0.882
WideDTA (27)	CNN	CNN	0.262	0.886
DeepCDA (30)	CNN+LSTM	CNN+LSTM	0.248	0.891
DeepGLSTM (33)	GCN	BiLSTM	0.232	0.895
DeepMHADTA (31)	CNN+ Attention	CNN+ Attention	0.244	0.895
GraphDTA-GCN (20)	GCN	CNN	0.254	0.880
GraphDTA-GAT-GCN (20)	GAT-GCN	CNN	0.245	0.881
GraphDTA-GAT (20)	GAT	CNN	0.232	0.892
GraphDTA-GIN (20)	GIN	CNN	0.229	0.893
AttentionDTA (28)	CNN+ Attention	CNN+ Attention	0.216	0.893
GDilatedDTA (36)	McGEN+ MLRCN	BiLSTM	0.237	0.885
KC-DTA (35)	GCN	2DCNN+3DCNN	0.241	0.884
KC-DTA (35)	GAT	2DCNN+3DCNN	0.240	0.886
KC-DTA (35)	GIN	2DCNN+3DCNN	0.235	0.888
MGraphDTA (34)	Multi Scale Deep GNN	Multi Scale CNN	0.207	0.900
DGraphDTA (37)	GCN	GCN	0.202	0.904
WGNNDTA-GCN (39)	GCN	GCN	0.208	0.902
WGNNDTA-GAT (39)	GAT	GAT	0.208	0.898
HiSIF-DTA(Top-Down) (38)	GCN	GCN	0.193	0.907
HiSIF-DTA(Bottom-Up) (38)	GCN	GCN	0.191	0.906
GTAMP-DTA (41)	GraphTransformer with Attention	GraphTransformer with Attention	0.177	0.923
TransVAE-DTA (40)	Transformer	VAE+Transformer	0.332	0.870
AttentionMGT-DTA (42)	GraphTransformer with Attention	3D-Graph(Transformer with Attention)+Sequence(Pre-trained embedding)	0.193	0.891
Proposed Model	**GAT+GIN**	**CNN**	**0.222**	**0.901**

TABLE 5. Performance of our model and the baselines on the Kiba dataset

Models	Drug	Protein	MSE	CI
KronRLS (25)	Smith-Waterman	PubChem-Sim	0.411	0.782
SimBoost (26)	Smith–Waterman	PubChem-Sim	0.222	0.836
DeepDTA (15)	CNN	CNN	0.194	0.863
GANsDTA (29)	GAN	GAN	0.224	0.866
DeepGS (32)	GAT+BiGRU	CNN	0.193	0.860

WideDTA (27)	CNN	CNN + PDM	0.179	0.875
DeepCDA (30)	CNN+LSTM	CNN+LSTM	0.176	0.889
DeepGLSTM (33)	GCN	BiLSTM	0.133	0.897
DeepMHADTA (31)	CNN+ Attention	CNN+ Attention	0.186	0.876
GraphDTA-GCN (20)	GCN	CNN	0.139	0.889
GraphDTA-GAT-GCN (20)	GAT-GCN	CNN	0.139	0.891
GraphDTA-GAT (20)	GAT	CNN	0.179	0.866
GraphDTA-GIN (20)	GIN	CNN	0.147	0.882
AttentionDTA (28)	CNN+ Attention	CNN+ Attention	0.155	0.882
GDilatedDTA (36)	McGEN+ MLRCN	BiLSTM	0.156	0.876
KC-DTA (35)	GCN	2DCNN+3DCNN	0.143	0.890
KC-DTA (35)	GAT	2DCNN+3DCNN	0.180	0.861
KC-DTA (35)	GIN	2DCNN+3DCNN	0.159	0.877
MGraphDTA (34)	Multi Scale Deep GNN	Multi Scale CNN	0.128	0.902
DGraphDTA (37)	GCN	GCN	0.126	0.904
WGNNDTA-GCN (39)	GCN	GCN	0.141	0.889
WGNNDTA-GAT (39)	GAT	GAT	0.130	0.900
HiSIF-DTA(Top-Down) (38)	GCN	GCN	0.120	0.904
HiSIF-DTA(Bottom-Up) (38)	GCN	GCN	0.121	0.904
GTAMP-DTA (41)	GraphTransformer with Attention	GraphTransformer with Attention	0.123	0.917
TransVAE-DTA (40)	Transformer	VAE+Transformer	0.253	0.822
AttentionMGT-DTA (42)	GraphTransformer with Attention	3D-Graph(Transformer with Attention)+Sequence(Pre-trained embedding)	0.140	0.893
Proposed Model	**GAT+GIN**	**CNN**	**0.144**	**0.883**

5. ABLATION EXPERIMENTS

To evaluate the contribution of different components of our proposed model, we conducted an ablation study by training and testing the model under different settings. Specifically, we considered three variations of our model:

- GAT + GIN: In this setting, we removed the cross-attention mechanism from our original model to evaluate its impact on the performance.
- GIN + Cross Attention: In this setting, we removed the GAT layer from our model and only used the GIN layers with cross attention layer to evaluate its impact on the performance.
- GAT + GIN + Cross Attention: This is our original proposed model, which utilizes both GAT and GIN layers, along with a cross-attention mechanism.

The results of our ablation study are summarized in Table 6.

TABLE 6. Results of ablation experiments on the Davis dataset

Model	MSE	CI
GAT + GIN(without cross Att)	0.225	0.891
GIN + Cross Att (without GAT)	0.232	0.888
GAT + GIN + Cross Attention(our proposed model)	**0.222**	**0.901**

As shown in Table 6 our proposed model (GAT+ GIN + Cross Attention) had first-rate performance in terms of MSE and CI. The consequences show that the mixture of GAT, GIN, and Cross Attention mechanisms effectively increases the accuracy and robustness of the model.

By removing the cross-attention mechanism from the proposed method (GAT +GIN), the performance reduced appreciably, with MSE increasing from 0.222 to 0.225 and CI lowering from 0.901 to 0.891. This indicates that the cross-attention mechanism plays an important role in predicting interactions between drugs and targets. Afterward while eliminating the GAT layer from our model (GIN+ Cross Attention), the performance considerably decreased, with MSE growing from 0.222 to 0.232 and CI reducing from 0.901 to 0. 888. Thus GAT layer is necessary for accurate predictions.

6. CONCLUSION

In this study, we proposed a new approach for drug-target affinity (DTA) prediction that leverages the power of GATs and GINs with cross attention mechanism. Our method captures both local and global structural information of the drug and protein by combining GATs and GINs, and learns the pairwise interactions between them through cross-attention.

The Cross-Attention layer, which facilitates the exchange of information between protein sequences and drugs was found to significantly enhance model performance. This combination offers interpretability and flexibility, enabling the identification of relevant substructures for binding affinity and handling variable-length inputs.

We evaluated our method on two widely used benchmark datasets, Davis and KIBA, consisting of 30056 and 118254 drug-target pairs, respectively. We as compared its performance with several strategies. The results confirmed that our proposed method handed present strategies, achieving a MSE of 0.222 and a CI of 0.901 on Davis, a MSE of 0.144, and a CI of 0.88 on KIBA. Our findings endorse that our approach has big potential for programs in drug discovery and layout, and will useful resource inside the identity of novel drug candidates and goals.

In summary, our studies present a new effective method for DTA prediction that mixes GATs and GINs with cross attention mechanism. Our method offers interpretability, flexibility, and advanced overall performance over present strategies, highlighting its capability for further applications in drug discovery.

7. REFERENCES

1. Momtahen S, Taajobian M, Jahanian A. Drug discovery acceleration using digital microfluidic biochip architecture and computer-aided-design flow. International Journal of Engineering, Transactions B: Applications. 2019;32(8):1169-76. 10.5829/ije.2019.32.08b.13
2. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Molecular pharmaceutics. 2016;13(7):2524-30. 10.1021/acs.molpharmaceut.6b00248
3. Dudley JT, Deshpande T, Butte AJ. Exploiting drug–disease relationships for computational drug repositioning. Briefings in bioinformatics. 2011;12(4):303-11. 10.1093/bib/bbr013
4. Mak K-K, Wong Y-H, Pichika MR. Artificial intelligence in drug discovery and development. Drug discovery and evaluation: safety and pharmacokinetic assays. 2024:1461-98. 10.1016/j.drudis.2020.10.010
5. Ghanbari Sorkhi A, Hashemi S, Yarmohammadi H, Iranpour Mobarakeh M. Prediction of Drug-Target Protein Interaction Based on the Minimization of Weighted Nuclear Norm and Similarity Graph between Drugs and Target Proteins. International Journal of Engineering, Transactions A: Basics. 2021;34(7):1736-42. 10.5829/ije.2021.34.07a.18
6. Zohrevand A, Imani Z, Ezoji M. Deep convolutional neural network for finger-knuckle-print recognition. International Journal of Engineering, Transactions A: Basics. 2021;34(7):1684-93. 10.5829/ije.2021.34.07a.12
7. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data. 2021;8:1-74. 10.1186/s40537-021-00444-8
8. Zeng X, Li S-J, Lv S-Q, Wen M-L, Li Y. A comprehensive review of the recent advances on predicting drug-target affinity based on deep learning. Frontiers in Pharmacology. 2024;15:1375522.
9. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. Briefings in bioinformatics. 2021;22(1):247-69. 10.1093/bib/bbz157
10. Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chemical science. 2018;9(24):5441-51. 10.1039/c8sc00148k
11. Sachdev K, Gupta MK. A comprehensive review of feature based methods for drug target interaction prediction. Journal of biomedical informatics. 2019;93:103159. 10.1016/j.jbi.2019.103159
12. Zhang Y, Hu Y, Han N, Yang A, Liu X, Cai H. A survey of drug-target interaction and affinity prediction methods via graph neural networks. Computers in Biology and Medicine. 2023;163:107136. 10.1016/j.compbiomed.2023.10713
13. Rifaioglu AS, Cetin Atalay R, Cansen Kahraman D, Doğan T, Martin M, Atalay V. MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. Bioinformatics. 2021;37(5):693-704. 10.1093/bioinformatics/btaa858
14. Shim J, Hong Z-Y, Sohn I, Hwang C. Prediction of drug–target binding affinity using similarity-based convolutional neural network. Scientific Reports. 2021;11(1):4416. 10.1038/s41598-021-83679-y
15. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics. 2018;34(17):i821-i9. 10.1093/bioinformatics/bty593
16. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of chemical information and computer sciences. 1988;28(1):31-6. 10.1021/ci00057a005
17. Zhang S, Jiang M, Wang S, Wang X, Wei Z, Li Z. SAG-DTA: prediction of drug–target affinity using self-attention graph network. International Journal of Molecular Sciences. 2021;22(16):8993. 10.3390/ijms22168993
18. He H, Chen G, Chen CY-C. NHGNN-DTA: a node-adaptive hybrid graph neural network for interpretable drug–target binding affinity prediction. Bioinformatics. 2023;39(6):btad355. 10.1093/bioinformatics/btad355
19. Wang H, Zhou G, Liu S, Jiang J-Y, Wang W. Drug-target interaction prediction with graph attention networks. arXiv preprint arXiv:210706099. 2021. 10.48550/arXiv.2107.06099
20. Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140-7. 10.1093/bioinformatics/btaa921
21. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. arXiv preprint arXiv:171010903. 2017. 10.48550/arXiv.1710.10903

22. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? arXiv preprint arXiv:181000826. 2018. 10.48550/arXiv.1810.00826
23. Hou R, Chang H, Ma B, Shan S, Chen X. Cross attention network for few-shot classification. Advances in neural information processing systems. 2019;32. 10.48550/arXiv.1910.07677
24. Ke L, Li X, Danelljan M, Tai Y-W, Tang C-K, Yu F. Prototypical cross-attention networks for multiple object tracking and segmentation. Advances in neural information processing systems. 2021;34:1192-203. 10.48550/arXiv.2106.11958
25. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug–target interaction predictions. Briefings in bioinformatics. 2015;16(2):325-37. 10.1093/bib/bbu010
26. He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. Journal of cheminformatics. 2017;9:1-14. 10.1186/s13321-017-0209-z
27. Öztürk H, Ozkirimli E, Özgür A. WideDTA: prediction of drug-target binding affinity. arXiv preprint arXiv:190204166. 2019. 10.48550/arXiv.1902.04166
28. Zhao Q, Xiao F, Yang M, Li Y, Wang J, editors. AttentionDTA: prediction of drug–target binding affinity using attention model. 2019 IEEE international conference on bioinformatics and biomedicine (BIBM); 2019: IEEE. 10.1109/BIBM47256.2019.8983125
29. Zhao L, Wang J, Pang L, Liu Y, Zhang J. GANsDTA: Predicting drug-target binding affinity using GANs. Frontiers in genetics. 2020;10:1243. 10.3389/fgene.2019.01243
30. Abbasi K, Razzaghi P, Poso A, Amanlou M, Ghasemi JB, Masoudi-Nejad A. DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. Bioinformatics. 2020;36(17):4633-42. 10.1093/bioinformatics/btaa544
31. Deng L, Zeng Y, Liu H, Liu Z, Liu X. DeepMHADTA: prediction of drug-target binding affinity using multi-head self-attention and convolutional neural network. Current Issues in Molecular Biology. 2022;44(5):2287-99. 10.3390/cimb44050155
32. Lin X, Zhao K, Xiao T, Quan Z, Wang Z-J, Yu PS. DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. ECAI 2020: IOS Press; 2020. p. 1301-8.
33. Mukherjee S, Ghosh M, Basuchowdhuri P, editors. DeepGLSTM: deep graph convolutional network and LSTM based approach for predicting drug-target binding affinity. Proceedings of the 2022 SIAM international conference on data mining (SDM); 2022: SIAM. 10.1137/1.9781611977172.82
34. Yang Z, Zhong W, Zhao L, Chen CY-C. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. Chemical science. 2022;13(3):816-33. https://doi.org/10.1039/d1sc05180f
35. Jiang M, Shao Y, Zhang Y, Zhou W, Pang S. A deep learning method for drug-target affinity prediction based on sequence interaction information mining. PeerJ. 2023;11:e16625. https://doi.org/10.7717/peerj.16625
36. Zhang L, Zeng W, Chen J, Chen J, Li K. GDilatedDTA: Graph dilation convolution strategy for drug target binding affinity prediction. Biomedical Signal Processing and Control. 2024;92:106110. https://doi.org/10.1016/j.bspc.2024.106110
37. Jiang M, Li Z, Zhang S, Wang S, Wang X, Yuan Q, et al. Drug–target affinity prediction using graph neural network and contact maps. RSC advances. 2020;10(35):20701-12. 10.1039/D0RA02297G
38. Bi X, Zhang S, Ma W, Jiang H, Wei Z. HiSIF-DTA: a hierarchical semantic information fusion framework for drug-target affinity prediction. IEEE Journal of Biomedical and Health Informatics. 2023. 10.1109/JBHI.2023.3334239
39. Jiang M, Wang S, Zhang S, Zhou W, Zhang Y, Li Z. Sequence-based drug-target affinity prediction using weighted graph neural networks. BMC genomics. 2022;23(1):449. https://doi.org/10.1186/s12864-022-08648-9
40. Zhou C, Li Z, Song J, Xiang W. TransVAE-DTA: Transformer and variational autoencoder network for drug-target binding affinity prediction. Computer Methods and Programs in Biomedicine. 2024;244:108003. 10.1016/j.cmpb.2023.108003
41. Tian C, Wang L, Cui Z, Wu H. GTAMP-DTA: Graph transformer combined with attention mechanism for drug-target binding affinity prediction. Computational Biology and Chemistry. 2024;108:107982. 10.1016/j.compbiolchem.2023.107982
42. Wu H, Liu J, Jiang T, Zou Q, Qi S, Cui Z, et al. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. Neural Networks. 2024;169:623-36. 10.1016/j.neunet.2023.11.018
43. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nature biotechnology. 2011;29(11):1046-51. 10.1038/nbt.1990
44. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. Journal of Chemical Information and Modeling. 2014;54(3):735-43. 10.1021/ci400709d
45. Landrum G. RDKit: Open-source cheminformatics. 2006. Google Scholar. 2006.
46. Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM. Unified rational protein engineering with sequence-based deep representation learning. Nature methods. 2019;16(12):1315-22. 10.1038/s41592-019-0598-1
47. Nitish S. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1. 10.1145/3065386
48. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6):84-90.
49. Vaswani A. Attention is all you need. Advances in Neural Information Processing Systems. 2017.
50. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data. Journal of computational chemistry. 2013;34(12):1071-82. 10.1002/jcc.23231
51. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32. 10.48550/arXiv.1912.01703

Persian Abstract

چکیده

در فرایند کشف دارو، تعاملات دارو–هدف (DTA) به عنوان یک مرحله حیاتی در نظر گرفته میشود، زیرا به شناسایی امیدوارکننده ترین نامزدهای دارویی در فرآیند توسعه دارو کمک میکند. از آنجا که ساختار و عملکرد مولکولهای دارو و هدف و همچنین تعاملات پیچیده و غیرخطی آنها باید در نظر گرفته شود، پیشبینی DTA یک فرایند چالش برانگیز است. هدف این مطالعه پیشنهاد یک چارچوب جدید پیشبینی DTA است که از نقاط قوت شبکه های توجه متقابل (CANs) با استفاده از شبکه های عصبی گراف (GNNs) بهره میبرد. استفاده از گرافها با استفاده از GNNs اطلاعات ساختاری سه بعدی آنها را حفظ میکند، اما توسط روشهای مبتنی بر توجه موجود به طور کامل ارائه وتفسیر نمیشوند. چارچوب ما از CAN برای گرفتن نمایندگی دقیقتر از جفت دارو–هدف با تحلیل چگونگی تعامل بخشهای مختلف یک مولکول دارو با مناطق خاصی از پروتئین استفاده میکند. ما از GIN و GAT در یک معماری متوالی برای گرفتن اطلاعات ساختاری محلی و جهانی مولکولهای گراف دارو استفاده کردیم. عملکرد روش پیشنهادی را بر روی دو مجموعه داده مرجع، و KIBA، ارزیابی کردیم. عملکرد امیدوارکننده است و از بسیاری روشهای پیشرفته از نظر میانگین مربعات خطا (MSE) و شاخص همبستگی (CI) پیشی میگیرد. به طور خاص، برای مجموعه داده DAVIS ، MSE برابر با ۰.۲۲۲ و CI برابر با ۰.۹۰۱ به دست آوردیم، در حالی که برای KIBA، MSE برابر با ۰.۱٤٤ و CI برابر با ۰.۸۸۳ به دست آوردیم. روش ما قابلیت تفسیر و خاصیت تحلیل تعامل را افزایش میدهد، بینش عمیقتری در فرآیند کشف دارو فراهم میکند و توضیحات ارزشمندی برای DTA پیشبینی شده ارائه میدهد.