Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 6, No. 3, 2026

Drug-Target Interaction Prediction Based on Deep Feature Fusion

Zhiyuan Xu *

College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China

Abstract: Drug-target interaction prediction plays an important role in drug discovery and drug repositioning. Traditional experimental screening methods are expensive and time-consuming, while existing computational approaches still have limitations in molecular representation and feature fusion. To address these problems, this paper proposes a drug-target interaction prediction model based on deep feature fusion. In the GCT-DTI, a Graph Isomorphism Network is used to capture the topological structure information of drug molecules, and Morgan fingerprints are introduced to supplement global chemical features. For target proteins, a Transformer encoder is employed to learn long-range dependency information from amino acid sequences. In addition, a cross-attention mechanism is designed to enhance the interaction between drug features and protein features. Experiments on the Human and C.elegans benchmark datasets demonstrate that the GCT-DTI achieves good performance in terms of AUC, Precision, and Recall. The results show that deep feature fusion can effectively improve the prediction accuracy of drug-target interactions and provide useful support for computer-aided drug discovery.

Keywords: Drug-target interaction, deep learning, feature fusion, graph neural network, Transformer.

1. Introduction

Drug discovery is a complex process with high cost, long development cycles, and high risk [1, 2]. In recent years, identifying potential drug-target interactions has become an important task in early-stage drug development. Accurate drug-target interaction prediction can help researchers quickly screen candidate compounds, reduce experimental costs, and improve the efficiency of new drug discovery.

Traditional methods for drug-target interaction prediction mainly include molecular docking, similarity-based methods, and machine learning methods [3]. Molecular docking can provide intuitive information about binding modes, but it usually depends on accurate three-dimensional structures of proteins and requires high computational cost. Similarity-based methods are efficient, but they rely heavily on known interaction networks and often suffer from the cold-start problem. Traditional machine learning methods can model the relationship between drugs and targets to some extent, but their performance is strongly affected by handcrafted features.

With the rapid development of artificial intelligence, deep learning has shown strong capability in biological data analysis [4-7]. Existing deep learning-based methods have achieved promising results in drug-target interaction prediction. However, they still face several challenges. First, some models cannot fully capture both local structural information and global chemical information of drug molecules. Second, conventional sequence models may fail to effectively model long-range dependencies in protein sequences. Third, simple feature concatenation is not sufficient to describe the complex interaction mechanism between drugs and target proteins.

To overcome these limitations, this paper proposes a drug-target interaction prediction method based on deep feature fusion. In the proposed model, drug molecules are represented from both graph structure and fingerprint views. Protein sequences are encoded by a Transformer-based module [8-10]. Then, a cross-attention mechanism is introduced to fuse drug and protein features more effectively. Experimental results on public benchmark datasets demonstrate the effectiveness of the GCT-DTI.

2. Proposed Method

2.1. Overall Framework

The proposed model mainly consists of three modules: drug feature extraction, protein feature extraction, and feature fusion for interaction prediction. First, the drug is represented by both molecular graph information and Morgan fingerprint information. Second, the protein is represented by its amino acid sequence. Third, a cross-attention mechanism is used to fuse the drug and protein features. Finally, the fused feature vector is fed into a multilayer perceptron for binary classification.

2.2. Drug Feature Extraction

To comprehensively represent drug molecules, both local topology information and global chemical information are considered in this work.

For local topology representation, the SMILES string of each drug is converted into a molecular graph by using RDKit. In the molecular graph, atoms are treated as nodes and chemical bonds are treated as edges. Each atom is encoded with its chemical properties to form the node feature matrix. Then, a Graph Isomorphism Network (GIN) is used to extract graph-based features from the molecular graph [11]. Compared with conventional graph convolution methods, GIN has stronger ability to distinguish subtle topological differences among molecular structures. After several graph convolution layers, a global pooling operation is applied to obtain the graph-level representation of the drug.

For global chemical representation, Morgan fingerprints are used to describe important substructures and functional groups of drug molecules. In this study, the fingerprint radius is set to 2 and the fingerprint length is set to 1024. To further enhance the representation ability of the fingerprint features, a self-attention mechanism is introduced to model the dependency among fingerprint components. The enhanced

fingerprint feature is then projected to the same feature space as the graph feature.

Finally, the graph feature and the fingerprint feature are concatenated to form the final drug representation. In this way, the model can simultaneously learn local structural characteristics and global chemical semantics of drug molecules.

2.3. Protein Feature Extraction

Protein sequences contain rich biological information, and long-range dependency among amino acid residues is important for drug-target interaction prediction. Therefore, this paper adopts a Transformer-based encoder to extract protein features.

First, the amino acid sequence is segmented by an N-gram strategy, where the window size is set to 3. This operation can preserve local contextual information of residues and transform the original sequence into a token sequence. Then, each token is embedded into a low-dimensional vector space. Positional encoding is added to the embedding vectors so that the model can capture the sequential order of amino acids.

After that, the embedded sequence is fed into the Transformer encoder. Through the multi-head self-attention mechanism, the Transformer can learn contextual dependencies among different positions in the protein sequence, including long-distance relationships. This helps the model better capture important biological patterns in protein sequences. The output of the Transformer encoder is used as the final protein representation.

2.4. Feature Fusion

After obtaining the drug feature and protein feature, the next key problem is how to effectively model the interaction between them. Traditional concatenation-based methods are usually too simple to capture complex cross-modal relationships. Therefore, this paper introduces a cross-attention mechanism for feature fusion.

In the cross-attention module, the drug feature is used as the query, while the protein sequence representation is used as the key and value. By calculating attention weights, the model can identify which parts of the protein sequence are more relevant to the given drug. Then, a weighted protein context vector is generated according to the attention scores. This process allows the model to focus on important interaction-related regions.

The final fused feature is obtained by concatenating the original drug feature and the protein context feature. Compared with direct concatenation, the proposed feature fusion strategy can better model the potential interaction relationship between drugs and proteins.

2.5. Interaction Prediction

The fused feature vector is input into a multilayer perceptron for binary classification. A sigmoid activation function is used in the output layer to generate the probability that a drug-target pair has an interaction. Since drug-target interaction prediction is a binary classification task, the binary cross-entropy loss is used to optimize the model parameters.

3. Experiments and Results

3.1. Experimental Settings

The proposed model is implemented under a unified experimental environment. RDKit is used for molecular graph construction and fingerprint generation. Pandas and NumPy are used for data preprocessing. The model is trained on a server equipped with Intel Xeon CPU and NVIDIA GeForce RTX 4090 GPU.

The main hyperparameters of the model are as follows. The hidden embedding dimension is set to 64. The batch size is 16. The initial learning rate is 5e-4. The learning rate decays by a factor of 0.5 every 10 epochs. The maximum training epoch is 100. The N-gram window size of protein sequences is 3. The number of GIN layers is 3. The radius of Morgan fingerprint is 2, and the fingerprint length is 1024. The dropout rates of feature extraction layer and classification layer are 0.1 and 0.5, respectively.

3.2. Datasets

To evaluate the effectiveness of the GCT-DTI, two benchmark datasets are used in this study: Human and C.elegans. These two datasets are widely used in drug-target interaction prediction tasks.

The Human dataset contains 1,052 unique compounds and 852 unique target proteins, including 6,212 drug-target pairs. Among them, 3,369 are positive samples and 2,843 are negative samples.

The C.elegans dataset contains 1,434 unique compounds and 2,504 unique target proteins, including 7,511 drug-target pairs. Among them, 4,000 are positive samples and 3,511 are negative samples.

Table 1. Statistics of benchmark datasets

Dataset	Compounds	Target Proteins	Interactions	Positive Samples	Negative Samples
Human	1052	852	6212	3369	2843
C.elegans	1434	2504	7511	4000	3511

3.3. Evaluation Metrics

Since drug-target interaction prediction is a binary classification task, three evaluation metrics are used in this paper: AUC, Precision, and Recall.

AUC reflects the overall ability of the model to distinguish positive and negative samples. Precision measures the proportion of true positive samples among all predicted positive samples. Recall measures the proportion of correctly identified positive samples among all actual positive samples. Higher values of these metrics indicate better model performance.

3.4. Ablation Study

To verify the contribution of each component in the GCT-DTI, ablation experiments are conducted on the Human and C.elegans datasets. Several variants are designed by removing or replacing important modules, including GIN, Transformer, Morgan fingerprint branch, self-attention module, and cross-attention fusion module.

Table 2. Ablation results on the Human dataset

Method	AUC	Recall	Precision
w/o attention	0.971	0.942	0.941
w/o morgan	0.968	0.938	0.935
w/o Transformer	0.965	0.935	0.928
w/o Cross-Attention	0.958	0.925	0.921
w/o GIN	0.970	0.940	0.939
GCT-DTI	0.974	0.945	0.948

Table 3. Ablation results on the C.elegans dataset

Method	AUC	Recall	Precision
w/o attention	0.981	0.955	0.949
w/o morgan	0.981	0.949	0.942
w/o Transformer	0.975	0.945	0.937
w/o Cross-Attention	0.969	0.935	0.925
w/o GIN	0.979	0.952	0.938
GCT-DTI	0.983	0.958	0.952

The ablation results show that the complete model achieves the best performance on both datasets. When the cross-attention module is removed, the performance decreases most obviously, which indicates that the interaction modeling between drug features and protein features is very important. Replacing Transformer with a simpler sequence model also causes a decline in performance, showing that long-range dependency modeling is beneficial for protein sequence representation. In addition, removing the Morgan fingerprint branch leads to lower results, which proves that global chemical information can effectively complement graph-based features. These results demonstrate the effectiveness of deep feature fusion.

3.5. Comparative Results

To further evaluate the GCT-DTI, it is compared with several classical machine learning and deep learning methods on the Human and C.elegans datasets [12].

Table 4. Comparison results on the Human dataset

Method	AUC	Precision	Recall
RF	0.940	0.897	0.861
KNN	0.860	0.927	0.798
DrugVQA (VQA-seq)	0.960	0.882	0.912
GraphDTA	0.956	0.862	0.928
GCN	0.910	0.966	0.969
CPI-GNN	0.970	0.918	0.923
DeepConv-DTI	0.970	0.942	0.953
GCT-DTI	0.974	0.948	0.945

Table 5. Comparison results on the C.elegans dataset

Method	AUC	Precision	Recall
RF	0.902	0.821	0.844
L2	0.892	0.890	0.877
SVM	0.894	0.785	0.818
GraphDTA	0.974	0.974	0.912
GCN	0.975	0.921	0.927
CPI-GNN	0.978	0.938	0.929
DeepConv-DTI	0.979	0.947	0.950
GCT-DTI	0.983	0.958	0.952

From the comparison results, it can be seen that the GCT-DTI achieves the best AUC on both Human and C.elegans datasets. On the Human dataset, the GCT-DTI reaches an AUC of 0.974, which is higher than the competing methods. On the C.elegans dataset, the GCT-DTI obtains an AUC of 0.983 and also achieves competitive Precision and Recall values. These results indicate that the proposed deep feature fusion model has stronger capability in identifying potential drug-target interactions.

The good performance of the GCT-DTI can be mainly attributed to three aspects. First, the GIN module improves the representation ability of molecular graphs and captures subtle topological differences among drugs. Second, the Transformer encoder effectively models long-range dependencies in protein sequences. Third, the cross-attention mechanism enables more effective fusion of drug and protein features. Therefore, the GCT-DTI can better learn the complex interaction patterns between drugs and target proteins.

4. Conclusion

In this paper, a drug-target interaction prediction method based on deep feature fusion is proposed. The method combines GIN-based molecular graph learning, Morgan fingerprint representation, Transformer-based protein sequence encoding, and cross-attention-based feature fusion. By integrating local structural information, global chemical information, and protein sequence semantics, the GCT-DTI can effectively improve the prediction accuracy of drug-target interactions.

Experiments on the Human and C.elegans benchmark datasets show that the GCT-DTI achieves good performance in terms of AUC, Precision, and Recall. The ablation study further verifies the effectiveness of each important module. The results demonstrate that deep feature fusion is a useful strategy for drug-target interaction prediction.

In future work, more biological information such as three-dimensional protein structure and pretrained molecular representation can be introduced to further improve the model performance and interpretability.

References

[1] Hughes J P, Rees S, Kalindjian S B, et al. Principles of early drug discovery. British Journal of Pharmacology, 2011, 162(6): 1239-1249.

[2] Berdigaliyev N, Aljofan M. An overview of drug discovery and development. Future Medicinal Chemistry, 2020, 12(10): 939-947.

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

[4] Lee I, Keum J, Nam H. DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. PLoS Computational Biology, 2019, 15(6): e1007129.

[5] Zheng S, Li Y, Chen S, et al. Predicting drug-protein interaction using quasi-visual question answering system. Nature Machine Intelligence, 2020, 2(2): 134-140.

[6] Nguyen T, Le H, Quinn T P, et al. GraphDTA: predicting drug-target binding affinity with graph neural networks. Bioinformatics, 2021, 37(8): 1140-1147.

[7] Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics, 2019, 35(2): 309-318.

[8] Wang X, Xia Z, Feng R, et al. SMFF-DTA: using a sequential multi-feature fusion method with multiple attention mechanisms to predict drug-target binding affinity. BMC Biology, 2025, 23(1): 120.

[9] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.

[10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.

[11] Wang L, You Z H, Chen X, et al. A computational-based method for predicting drug-target interactions by using stacked autoencoder deep neural network. Journal of Computational Biology, 2018, 25(3): 361-373.

[12] Zhao W, Yu Y, Liu G, et al. MSI-DTI: predicting drug-target interaction based on multi-source information and multi-head self-attention. Briefings in Bioinformatics, 2024, 25(3): bbae238.