Computers and Artificial Intelligence；Vol.2, No.3; 2025
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.25v2n3.2028
Published by: Zeus Press

CoA-DTI: A Drug–Target Interaction Prediction Model Based on Co-Attention

Bolun Qi*

School of Information, Guizhou University of Finance and Economics, Guiyang, China

**Corresponding author: Bolun Qi*

Abstract

Accurate prediction of drug–target interactions (DTIs) is a fundamental task in AI-driven drug discovery and repositioning. While traditional experimental methods are reliable, they are often time-consuming and expensive, limiting their scalability. Recent computational approaches have shown promise but still face challenges in extracting informative representations and modeling cross-modal interactions between drugs and proteins. In this work, we propose CoA-DTI, a novel deep learning framework that integrates Graph Convolutional Networks (GCNs), Convolutional Neural Networks (CNNs), and a Co-Attention mechanism for DTI prediction. Drug molecules are encoded as molecular graphs and processed through multi-layer GCNs to capture topological and atom-level features. Protein sequences are encoded via CNNs to learn local biochemical patterns. A co-attention module is then employed to model bidirectional interactions between the drug and protein representations, enabling fine-grained and context-aware feature fusion. Extensive experiments on the benchmark Davis dataset show that CoA-DTI achieves superior performance over competitive baselines in terms of accuracy and robustness. Ablation studies further validate the effectiveness of each module. This work demonstrates the potential of integrating graph-based learning and cross-modal attention in bioinformatics applications, offering a generalizable and interpretable approach for interaction prediction.

Keywords

drug–target interaction, graph neural networks, deep learning, co-attention mechanism, cross-modal feature fusion

1. Introduction

Drug–target interaction prediction is a fundamental problem in drug discovery and drug repurposing. Accurately identifying potential binding relationships between drug compounds and target proteins can significantly reduce development costs and time, while improving the success rate of novel therapeutics. Traditional experimental approaches, such as high-throughput screening (HTS), although reliable, are often time-consuming, expensive, and limited in scalability (Lamb et al., 2006; MacArron et al., 2011). Consequently, computational DTI prediction has emerged as a promising alternative, aiming to efficiently and accurately identify candidate drug–target pairs through algorithmic modeling (Yamanishi et al., 2008).

With the rapid advancement of artificial intelligence, DTI prediction methods have evolved substantially. Early studies primarily relied on handcrafted features, such as molecular fingerprints and sequence similarity, in combination with classical machine learning models—e.g., support vector machines, logistic regression,

and random forests (Breiman, 2001; Cortes & Vapnik, 1995; Fan et al., 2008). However, these approaches were limited in their ability to capture complex structural and functional patterns (Chen et al., 2012). More recently, graph neural networks have been widely adopted to model the topological structure of drug molecules represented as graphs, enabling effective extraction of structural features (Gilmer et al., 2017). In parallel, deep learning models such as convolutional neural networks have shown strong performance in learning protein sequence representations by capturing local motifs and functional domains (Li et al., 2024). Despite these improvements, most existing methods still struggle to effectively integrate and model interactions across heterogeneous modalities.

One of the key challenges in DTI prediction lies in fully leveraging the complementary information between drug structures and protein sequences and modeling their intricate cross-modal interactions. Conventional fusion strategies, such as simple concatenation or weighted averaging, often overlook the dynamic dependencies between modalities and fail to capture the underlying interaction mechanisms. The recent development of attention mechanisms offers a new paradigm for multimodal learning, particularly co-attention mechanisms, which jointly model inter-modal dependencies and facilitate precise alignment and interaction modeling between features from different sources (Lu et al., 2016).

In this work, we propose CoAG-DTI, a novel DTI prediction model that integrates GCNs with a co-attention-based fusion mechanism. Specifically, drug molecules are encoded through a multi-layer GCN to extract structural representations, while protein sequences are processed by a CNN to learn biologically meaningful sequence patterns. A co-attention module is then introduced to capture fine-grained, bidirectional interactions between drug and protein embeddings, enabling deep cross-modal feature fusion and enhancing predictive performance. We conduct extensive experiments on the standard Davis dataset, demonstrating that CoAG-DTI outperforms existing state-of-the-art methods in both accuracy and generalization. Ablation studies further confirm the effectiveness of the co-attention mechanism and the proposed fusion strategy.

The main contributions of this work are summarized as follows:We design a dual-modality feature extraction architecture that combines GNNs and CNNs to effectively encode structural and sequential features of drugs and proteins, respectively. We introduce a co-attention mechanism to enhance cross-modal interaction modeling and improve feature fusion quality. We perform comprehensive evaluation and ablation analysis on a public benchmark dataset, verifying the effectiveness and superiority of our model architecture.

2. Methods

2.1 Model Overview

To effectively model the complex interactions between drug molecules and target proteins, we propose a deep learning framework named CoAG-DTI, which integrates graph neural networks with a co-attention mechanism, as illustrated in Figure 1. The overall architecture consists of three core modules:

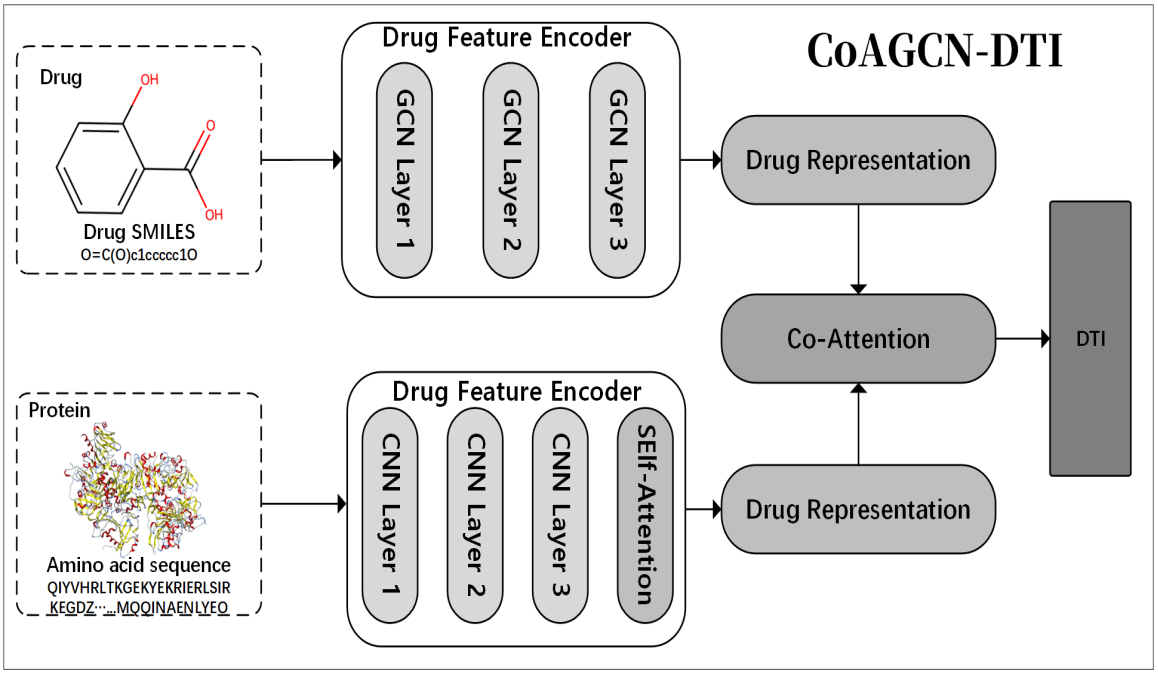
First, the drug feature encoder transforms each SMILES string into a molecular graph. A multi-layer graph convolutional network is then applied to extract both topological and atom-level features, capturing local and global structural information of the molecule.

Second, the protein feature encoder processes the primary amino acid sequence of the target protein using a multi-layer convolutional neural network, which learns local sequence patterns and structural motifs to produce biologically meaningful representations.

On top of these modality-specific encoders, we incorporate a co-attention mechanism that establishes a bidirectional attention mapping between the drug and protein representation spaces. This design enables the model to identify potential regions of interaction and facilitates deep cross-modal feature fusion.

Finally, the fused representation is fed into a fully connected layer to predict the probability of interaction between the drug and the target protein. This architecture maintains strong expressive capacity while enhancing the model's ability to capture fine-grained, functionally relevant interactions, ultimately improving the accuracy and interpretability of DTI prediction.

Figure 1: Overall architecture of the CoAG-DTI model



2.2 Drug Feature Encoder

In the CoAG-DTI model, drug molecules are first represented using the SMILES notation. Each SMILES string is parsed into a molecular graph $G = (V, E)$, where V denotes the set of atoms and E represents the chemical bonds between them. Each atom $v \in V$ is associated with a 75-dimensional initial feature vector x_v, encoding atom-level descriptors such as element type, degree, formal charge, and aromaticity.

To unify the representation dimension, a linear transformation is applied to project the input features into a shared embedding space:

$$\mathbf{h}_v^{(0)} = \mathbf{W}_0 \cdot \mathbf{x}_v$$

The transformed node embeddings are then passed through a stack of three GCN layers to extract local and global structural information. The propagation rule for the l-th GCN layer is defined as:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops added, $\tilde{D}$ is the corresponding degree matrix, $H^{(l)}$ is the node embedding matrix at layer l, $W^{(l)}$ is the learnable weight matrix, and $\sigma(\cdot)$ is a nonlinear activation function such as ReLU.

After the final GCN layer, the output node features are reshaped into a three-dimensional tensor of shape $[B, N, d]$, where B is the batch size, N is the maximum number of atoms per molecule, and d is the hidden dimension of the last GCN layer. This final representation, denoted as H_d, serves as the drug feature embedding and is used in the co-attention module for cross-modal interaction with protein features.

2.3 Protein Feature Encoder

To effectively capture biologically meaningful information from protein primary sequences, CoAG-DTI adopts a hybrid encoding strategy that combines CNN with Self-Attention .

Given a protein sequence $S = \{s_1, s_2, \ldots, s_L\}$, where each residue $s_i \in \mathcal{A}$ and $\mathcal{A}$ is the amino acid alphabet, we first embed each residue into a continuous vector space using an embedding layer, resulting in an input tensor of shape $\mathbb{R}^{L \times d_e}$, where d_e is the embedding dimension.

The embedded sequence is then passed through a series of three 1D convolutional layers, each followed by batch normalization and ReLU activation. The l-th convolutional layer operates as follows:

$$\mathrm{F}^{(\mathrm{l})} = \mathrm{ReLU}\left(\mathrm{BN}^{(\mathrm{l})}\left(\mathrm{Conv1D}^{(\mathrm{l})}(\mathrm{F}^{(\mathrm{l}-1)})\right)\right)$$

where $\mathrm{F}^{(0)}$ is the input embedding, and $\mathrm{BN}^{(\mathrm{l})}$ denotes batch normalization. The three convolutional layers use kernel sizes of 3, 5, and 7 to capture local sequential patterns at multiple scales.

After convolution, the resulting feature map is transposed into shape $[\mathrm{L}', \mathrm{B}, \mathrm{C}]$, where B is the batch size and C is the channel dimension. This tensor is then processed by a multi-head self-attention mechanism to model long-range dependencies between residues. Given the query Q, key K, and value V matrices, the attention output is computed as:

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

Multi-head attention is applied in parallel across multiple subspaces, and the outputs are concatenated and projected to obtain the final attention-enhanced sequence representation.

Finally, the output tensor is transposed and reshaped into shape $[\mathrm{B}, \mathrm{L}', \mathrm{d_p}]$, where $\mathrm{d_p}$ is the number of output channels. This representation $\mathrm{H_p}$ serves as the protein feature embedding, which is subsequently fed into the co-attention fusion module for interaction modeling with drug features.

2.4 Co-Attention Fusion Module

To model the fine-grained interactions between drugs and target proteins, we introduce a Co-Attention-based feature fusion module. This module takes as input the encoded drug representation $\mathrm{H_d} \in \mathbb{R}^{\mathrm{B}\times\mathrm{N}\times\mathrm{d}}$ and protein representation $\mathrm{H_p} \in \mathbb{R}^{\mathrm{B}\times\mathrm{L}\prime\times\mathrm{d_p}}$, which are obtained from the previous graph-based and sequence-based encoders, respectively. For notational clarity and to align with standard attention formulations, we denote $\mathrm{H_d}$ as V (value modality) and $\mathrm{H_p}$ as Q (query modality) in this section.

We first map both representations into a shared latent space with dimensionality d_h using linear transformations:

$$\widetilde{\mathbf{V}} = \mathbf{W}_v\mathbf{V}, \quad \widetilde{\mathbf{Q}} = \mathbf{W}_q\mathbf{Q}$$

where $\mathrm{W_v} \in \mathbb{R}^{\mathrm{d}\times\mathrm{d_h}}$ and $\mathrm{W_q} \in \mathbb{R}^{\mathrm{d_p}\times\mathrm{d_h}}$ are learnable projection matrices.

Next, we apply bi-directional multi-head attention to capture cross-modal dependencies:

$$\mathbf{A}_d = \mathrm{MultiHeadAttn}(\widetilde{\mathbf{V}}, \widetilde{\mathbf{Q}}, \widetilde{\mathbf{Q}}),\ \mathbf{A}_p = \mathrm{MultiHeadAttn}(\widetilde{\mathbf{Q}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{V}})$$

where $\mathrm{MultiHeadAttn}(\mathrm{Q}, \mathrm{K}, \mathrm{V})$ is the standard scaled dot-product attention mechanism applied across multiple heads.

To enhance feature stability and retain residual information, we apply dropout and layer normalization:

$$\mathbf{V}^{\mathrm{fused}} = \mathrm{LayerNorm}(\widetilde{\mathbf{V}} + \mathrm{Dropout}(\mathbf{A}_d)), \quad \mathbf{Q}^{\mathrm{fused}} = \mathrm{LayerNorm}(\widetilde{\mathbf{Q}} + \mathrm{Dropout}(\mathbf{A}_p))$$

Then, the fused features are globally aggregated via mean pooling over the sequence length:

$$\mathbf{v}_{\mathrm{agg}} = \mathrm{MeanPool}(\mathbf{V}^{\mathrm{fused}}), \quad \mathbf{q}_{\mathrm{agg}} = \mathrm{MeanPool}(\mathbf{Q}^{\mathrm{fused}})$$

Finally, the aggregated drug–protein representation is formed by concatenating both vectors:

$$\mathbf{z} = [\mathbf{v}_{\mathrm{agg}} \| \mathbf{q}_{\mathrm{agg}}]$$

where $\mathrm{z} \in \mathbb{R}^{2\mathrm{d_h}}$ is the fused cross-modal feature vector used for downstream interaction prediction.

3. Experimental Setup and Results

3.1 Dataset

We evaluate our model using the Davis dataset (Davis et al., 2011), which contains a total of 30,056 drug–target interaction pairs, involving 68 drug compounds and 442 target proteins. Among all samples, 9,533 are labeled as positive interactions and 20,523 as negative. This dataset covers more than 80% of the human catalytic protein kinase family, making it highly representative and valuable for research related to kinase-targeted drug development.

3.2 Baseline Models

We compare our method against a variety of baseline models, including both traditional machine learning and deep learning approaches.

Traditional models: Random Forest (RF) (Breiman, 2001), Support Vector Machine (SVM) (Cortes & Vapnik, 1995), and Naive Bayes (NB) (Bhuvaneswari & Kalaiselvi, 2012). These methods have been widely applied to various prediction tasks and are commonly used as performance benchmarks in DTI prediction.

Deep learning models:

NRLMFβ (Ban et al., 2019): An enhanced version of NRLMF that introduces a Beta-distribution scoring scheme to better fit low-resource interaction pairs.

HyperAttentionDTI (Zhao et al., 2022): Combines 1D convolutional layers with higher-order attention modules to capture spatial and channel-wise dependencies between drug and protein features.

EmbedDTI (Jin et al., 2021): Integrates protein language model embeddings with graph attention networks to learn hierarchical molecular representations.

GIFDTI (Zhao et al., 2023): Employs CNNFormer to extract both local and global sequence features and fuses them with global drug representations.

TransformerCPI (Chen et al., 2020): Uses a sequence-based Transformer architecture with Word2Vec and GCN encoders for compound and protein representation.

McANet (Bian et al., 2023): Proposes a shared-weight multi-head attention network optimized with PolyLoss to address data imbalance and overfitting.

GraphormerDTI (Gao et al., 2024): Leverages a graph-structured Transformer to model molecular topology for improved interaction prediction.

3.3 Evaluation Metrics

To scientifically and comprehensively evaluate the overall performance of drug–target interaction prediction models, this study adopts five key evaluation metrics. Accuracy measures the proportion of correctly predicted samples out of the total samples, reflecting the model's general correctness. Recall assesses the model's ability to correctly identify true positive cases, indicating its sensitivity in recognizing actual interactions. The F1 score, calculated as the harmonic mean of precision and recall, provides a balanced assessment that is particularly useful when dealing with imbalanced datasets where positive and negative samples are unevenly distributed. The area under the receiver operating characteristic curve (AUROC) quantifies the model's ability to distinguish between positive and negative samples across all classification thresholds by integrating the true positive rate against the false positive rate, with values closer to 1 indicating better discrimination. Additionally, the area under the precision-recall curve (AUPRC) emphasizes the quality of positive sample predictions by integrating precision and recall over varying thresholds, making it especially sensitive to datasets with class imbalance. Importantly, AUROC and AUPRC are threshold-independent metrics that comprehensively reflect the model's robustness and stability across different decision boundaries, whereas accuracy, recall, and F1 score are threshold-dependent and directly demonstrate the model's practical performance at specific classification thresholds. This multidimensional evaluation framework not only facilitates a fair and thorough comparison of different models' strengths and weaknesses but also enables

deeper insights into their behavior in terms of sensitivity, specificity, and other critical characteristics relevant to drug–target interaction prediction tasks.

3.4 Comparative Experiments

Table 1: Results of the models on the Davis dataset and their comparison with the latest methods. The best results are highlighted in bold, while the second-best results are underlined.

Model	AUROC	AUPRC	Accuracy	Recall	F1
RF	0.8964	0.8304	0.8448	0.6501	0.7201
NB	0.7011	0.5843	0.6753	0.6170	0.5385
SVM	0.8785	0.8016	0.8213	0.5970	0.6724
NRLMFβ	0.9003	0.8274	0.8207	0.7022	0.7389
HyperAttentionDTI	0.9139	<u>0.8330</u>	0.8636	0.7585	0.7596
EmbedDTI	0.8359	0.6652	0.8591	0.7756	0.7721
GIFDTI	0.9082	0.8211	0.8503	0.7522	0.7403
TransformerCPI	0.8645	0.7470	0.8208	0.6272	0.6637
McANet	<u>0.9256</u>	0.8511	<u>0.8705</u>	0.7667	0.7710
GraphormerDTI	0.8866	0.8022	0.8640	<u>0.8176</u>	<u>0.7984</u>
CoA-DTI	0.9321	0.8078	0.8913	0.8399	0.8560

In the comparative experiments on the Davis dataset, the CoA-DTI model demonstrated significant advantages across multiple key evaluation metrics. Specifically, it achieved an AUROC of 0.9321, an accuracy of 0.8913, a recall of 0.8399, and an F1 score of 0.8560, outperforming eight mainstream methods including RF, NB, SVM, and HyperAttentionDTI. Notably, CoA-DTI obtained the best results in accuracy, recall, and F1 score.

In summary, this experiment systematically validates the comprehensive performance superiority of the CoA-DTI model in the drug–target interaction prediction task.

3.5 Ablation Experiments

To thoroughly assess the contributions of key components within the CoA-DTI model to its overall predictive performance, we conducted ablation experiments by creating two distinct model variants. These variants isolate the effects of specific modules to better understand their individual impact. The results are summarized in Table 2.

w/o GCN: In this variant, the drug encoder no longer utilizes the original three-layer graph convolutional network. Instead, it is replaced with a more generic graph neural network architecture. This modification tests the importance of the specific GCN design in capturing molecular graph features.

w/o Co-A: This variant removes the co-attention mechanism from the model, thereby eliminating the cross-modal interactions between drug and protein features. Instead, the drug and protein representations are simply concatenated before being fed into the prediction layer. This setup evaluates the effectiveness of the co-attention module in enhancing feature fusion and interaction learning.

Table 2: Ablation study results — performance comparison across different model variants.

Model	AUROC	AUPRC	Accuracy	Recall
CoA-DTI	0.9321	0.8078	0.8913	0.8399
w/o GCN	0.9263	0.7786	0.7878	0.7804
w/o Co-A	0.9162	0.7943	0.7662	0.7585

The results of the ablation study clearly show that the CoA-DTI model achieves its best overall performance with the complete architecture, indicating that each designed module contributes positively to the task. When the GCN module in the drug encoder is removed and replaced with a generic GNN for drug representation, the model's performance declines to varying degrees across all metrics, demonstrating that the multi-layer GCN more effectively captures molecular graph structural information and enhances drug representation capability. Furthermore, when the co-attention mechanism is removed and drug and protein features are simply concatenated for prediction, the model's performance further deteriorates, especially in its ability to correctly identify positive samples. This indicates that the lack of effective cross-modal interaction weakens the model's capacity to capture underlying interaction relationships. Overall, both the GCN module and the co-attention

mechanism play crucial roles in improving feature extraction and drug–target interaction modeling, constituting core components that enable CoA-DTI to achieve superior performance.

4. Conclusion

In this work, we propose a novel deep learning model for drug–target interaction prediction, named CoA-DTI, which aims to enhance the fusion and interaction modeling of heterogeneous modality information. The model is built upon a graph attention-based drug encoder that effectively captures structural features of drug molecules through multi-layer graph convolution and self-attention mechanisms, while concurrently employing convolutional neural networks to extract local pattern information from protein sequences. To further strengthen cross-modal information exchange, a co-attention mechanism is introduced, enabling more interactive feature fusion between drugs and targets.

Experimental results on the Davis dataset demonstrate that CoA-DTI significantly outperforms several state-of-the-art methods, including GraphDTA, MolTrans, and HyperAttentionDTI, across multiple evaluation metrics such as AUROC, AUPRC, Accuracy, Recall, and F1 score. Ablation studies further confirm the critical roles of both the GCN module and the co-attention mechanism in boosting model performance.

In summary, CoA-DTI effectively models the complex nonlinear interactions between drugs and targets by deeply integrating graph neural networks with attention mechanisms, showing strong potential in DTI prediction tasks. Future work may explore incorporating protein structural information, knowledge graphs, or pretrained representations to further improve the model's generalization ability and biological interpretability.

References

Ban, T., Ohue, M., & Akiyama, Y. (2019). NRLMFβ: Beta-distribution-rescored neighborhood regularized logistic matrix factorization for improving the performance of drug–target interaction prediction. *Biochemistry and Biophysics Reports, 18*, Article 100615. https://doi.org/10.1016/J.BBREP.2019.01.008

Bhuvaneswari, R., & Kalaiselvi, K. (2012). Naive Bayesian classification approach in healthcare applications. *International Journal of Computer Science and Telecommunications, 3*(1), 106-112.

Bian, J., Zhang, X., Zhang, X., Xu, D., & Wang, G. (2023). MCANet: shared-weight-based MultiheadCrossAttention network for drug–target interaction prediction. *Briefings in Bioinformatics, 24*(2), Article bbad082. https://doi.org/10.1093/BIB/BBAD082

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

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. https://doi.org/10.1093/BIOINFORMATICS/BTAA524

Chen, X., Liu, M. X., & Yan, G. Y. (2012). Drug-target interaction prediction by random walk on the heterogeneous network. *Molecular BioSystems, 8*(7), 1970-1978. https://doi.org/10.1039/C2MB00002D

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. https://doi.org/10.1023/A:1022627411411

Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K., & Zarrinkar, P. P. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology, 29*(11), 1046-1051. https://doi.org/10.1038/NBT.1990

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *the Journal of machine Learning research, 9*, 1871-1874.

Gao, M., Zhang, D., Chen, Y., Zhang, Y., Wang, Z., Wang, X., Li, S., Guo, Y., Webb, G. I., Nguyen, A. T. N., May, L., & Song, J. (2024). GraphormerDTI: A graph transformer-based approach for drug-target interaction

prediction. *Computers in Biology and Medicine, 173*, Article 108339. https://doi.org/10.1016/J.COMPBIOMED.2024.108339

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). *Neural message passing for quantum chemistry* [Paper presentation]. Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia.

Jin, Y., Lu, J., Shi, R., & Yang, Y. (2021). EmbedDTI: Enhancing the molecular representations via sequence embedding and graph convolutional network for the prediction of drug-target interaction. *Biomolecules, 11*(12), Article 1783. https://doi.org/10.3390/BIOM11121783

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science, 313*(5795), 1929-1935. https://doi.org/10.1126/SCIENCE.1132939

Li, H., Zhang, W., Shang, Y., & Lu, W. (2024). *Collaborative prediction of drug-target interaction using sequence-based CNN and transformer* [Paper presentation]. Proceedings of the 2024 27th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2024, Tianjin, China.

Lu, J., Yang, J., Batra, D., & Parikh, D. (2016). *Hierarchical question-image co-attention for visual question answering* [Paper presentation]. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

MacArron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V. S., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U., & Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery, 10*(3), 188-195. https://doi.org/10.1038/NRD3368

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics, 24*(13), i232–i240. https://doi.org/10.1093/BIOINFORMATICS/BTN162

Zhao, Q., Duan, G., Zhao, H., Zheng, K., Li, Y., & Wang, J. (2023). GIFDTI: Prediction of drug-target interactions based on global molecular and intermolecular interaction representation learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20*(3), 1943-1952. https://doi.org/10.1109/TCBB.2022.3225423

Zhao, Q., Zhao, H., Zheng, K., & Wang, J. (2022). HyperAttentionDTI: improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics, 38*(3), 655-662. https://doi.org/10.1093/BIOINFORMATICS/BTAB715

Funding

This research was funded by Postgraduate Education and Teaching Reform Project of Guizhou Province in 2024 (No. 2024YJSKYJJ253).

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.