



Research article

Mutual-DTI: A mutual interaction feature-based neural network for drug-target protein interaction prediction

Jiahui Wen¹, Haitao Gan^{1,2,*}, Zhi Yang^{1,2}, Ran Zhou^{1,*}, Jing Zhao² and Zhiwei Ye¹

¹ School of Computer Science, Hubei University of Technology, Wuhan 430068, China

² State Key Laboratory of Biocatalysis and Enzyme Engineering, Hubei University, Wuhan 430062, China

* **Correspondence:** Email: htgan01@hbut.edu.cn, ranzhou@hbut.edu.cn.

Abstract: The prediction of drug-target protein interaction (DTI) is a crucial task in the development of new drugs in modern medicine. Accurately identifying DTI through computer simulations can significantly reduce development time and costs. In recent years, many sequence-based DTI prediction methods have been proposed, and introducing attention mechanisms has improved their forecasting performance. However, these methods have some shortcomings. For example, inappropriate dataset partitioning during data preprocessing can lead to overly optimistic prediction results. Additionally, only single non-covalent intermolecular interactions are considered in the DTI simulation, ignoring the complex interactions between their internal atoms and amino acids. In this paper, we propose a network model called Mutual-DTI that predicts DTI based on the interaction properties of sequences and a Transformer model. We use multi-head attention to extract the long-distance interdependent features of the sequence and introduce a module to extract the sequence's mutual interaction features in mining complex reaction processes of atoms and amino acids. We evaluate the experiments on two benchmark datasets, and the results show that Mutual-DTI outperforms the latest baseline significantly. In addition, we conduct ablation experiments on a label-inversion dataset that is split more rigorously. The results show that there is a significant improvement in the evaluation metrics after introducing the extracted sequence interaction feature module. This suggests that Mutual-DTI may contribute to modern medical drug development research. The experimental results show the effectiveness of our approach. The code for Mutual-DTI can be downloaded from <https://github.com/a610lab/Mutual-DTI>.

Keywords: drug-target protein interaction; drug discovery; mutual interaction features

1. Introduction

Predicting and identifying DTI is of great significance in medicine and biology [1–3]. Measuring the affinity with drugs and proteins in wet lab is the most expensive and time-consuming method. However, these experiments are often expensive and time-consuming due to the complexity of molecules [4]. Virtual screening (VS) through computation can significantly reduce costs. Structure-based VS and ligand-based VS, which are classical virtual screening methods, have achieved great success [5, 6]. Structure-based methods use 3-dimensional (3D) conformations of proteins and drugs for study of bioactivity. The ligand-based method is based on the assumption that similar molecules will interact with similar proteins [7]. However, the application of these methods is limited. For example, ligand-based VS methods perform poorly when the molecule has a few known binding proteins and structure-based VS methods cannot be executed when the 3D structure of a protein is unknown. Since the accurate reconstruction of protein structures is still to be developed, the construction of 3D-free DTI prediction methods has attracted increasing attention [8]. The machine learning approach considers the chemical space, genomic space and their interactions in a specific framework and formulates the DTI prediction as a classification problem following a feature-based and similarity-based approach. Similarity-based approaches rely on the assumption that drugs with similar structures should have similar effects, feature-based approaches construct a feature vector consisting of a combination of descriptors for the drug and the protein as model inputs. Bleakley et al. [9] proposed a new supervised method of inference to predict unknown drug-target interactions, which uses support vector machines as local classifiers. Since then, a variety of machine learning-based algorithms have been proposed that consider both composites and protein information in a unified model [10–19].

In recent years, the development of deep learning in drug discovery has been rapid. In comparison to traditional machine learning, end-to-end models eliminate the need to define and compute descriptors before modeling, providing different strategies and representations for proteins and drugs. Initially, manually specified descriptors were used to represent drugs and proteins, and a fully connected neural network (FCN) was designed to make predictions [20]. Since descriptors are designed from a single perspective and cannot be changed during the training process, descriptor-based approaches cannot extract task-relevant features. Therefore, many end-to-end models have been proposed. Lee et al. [21] proposed a model called DeepConv-DTI to predict DTI. The model uses convolution layer to extract local residue features of generalized proteins. Tsubaki et al. [22] used different models to represent drugs and proteins from the perspective of considering the structure of drugs as graph structures. Graph neural networks (GNN) were used to learn features from drug sequences, and convolutional neural network (CNN) was used to train protein sequences. In order to consider deeper features between molecules, Li et al. [23] proposed a multi-objective neural network (MONN), which introduced a gate recurrent unit (GRU) module to predict the affinity and can accurately determine the interaction and affinity between molecules. Zamora-Resendiz et al. [24] defined a new spatial graph convolutional network (GCN) architecture that employs graph reduction to reduce the number of training parameters and facilitate the abstraction of intermediate representations. Ryu et al. [25] combined GCN with an attention mechanism to enable GCN to identify atoms in different environments, which could extract better structural features related to a target molecular property such as solubility, polarity, synthetic accessibility and photovoltaic efficiency, compared to the vanilla GCN. Ru et al. [26] combined the ideas of adjacency and learning to rank to establish correlations between proteins and drugs using adjacency,

and predicted the binding affinity of drugs and proteins using learning to rank methods to input features into the classifier. Transformer is a model that uses a self-attention to improve the speed of model training, and has achieved great success in the field of natural language processing. Wang et al. [27] first extracted drug carriers by GNN, and then represented protein features by Transformer and CNN to obtain remote interaction information in proteins by a one-sided attention mechanism. Chen et al. [28] obtained the interaction features by Transformer decoder and proposed a more rigorous method for data set partitioning. Ren et al. [29] presented a deep learning framework based on multimodal representation and meta-path semantic analysis, which drugs and proteins are represented as multimodal data and the relationships between them are captured by meta-path semantic analysis. However, most of these methods only consider a single non-covalent interaction between drugs and proteins. In fact, there is much more than one interaction between drugs and proteins.

Inspired by the Transformer decoder, which is able to extract long-range interdependent features [28], this paper proposed a dual-pathway model for DTI prediction based on mutual reaction features, called Mutual-DTI. The transformer's decoder was modified to treat drugs and proteins as two distinct sequences. Additionally, a module was added to extract mutual features that enable learning of the complex interactions between atoms and amino acids. Figure 1 shows an overview of the entire network. The dual pathway approach has also been applied in other field. For example, dual attention matching (DAM) [30] was proposed to learn the global features from local features with self-attention but ignored the mutual influence information between local features of two modalities.

In this paper, we captured the spatial and other feature information of drugs with a GNN and represented the protein features with a gated convolutional network. The drug features and protein features were then input as two sequences into the Transformer decoder, which included the mutual feature module. Different from DAM, the mutual feature module simultaneously considered the local features of both drug molecules and proteins, which effectively extracted the interaction features between two sequences. Finally, the drug-protein feature vector was input into the fully connected layer for prediction. We expected the Mutual-DTI model to exhibit better performance and generality with the addition of the mutual feature module. To validate this, we evaluated it on two benchmark datasets and conducted ablation experiments on a more tightly delineated dataset [28]. The results showed that Mutual-DTI exhibited better performance. We further visualized the attention scores obtained by Mutual-DTI learning, and the results showed that the mutual feature module of Mutual-DTI helped to reduce the search space of binding points.

2. Materials and methods

2.1. Drug model

GNN aggregates operations to extract node features in a graph. We represent drugs using a graph structure where nodes represent atoms like carbon and hydrogen, and edges represent chemical bonds like single and double bonds. We use the RDKit Python library* to convert a simplified molecular input line entry system (SMILES) string into a two-dimensional drug molecular graph.

We define a drug graph by $G = \{V, E\}$, where V is the set of atomic nodes and E is the set of chemical bond edges. Considering the small number of atomic and chemical bond types, we perform a local breadth-first search for the nodes in the graph, where the search depth r is equal to the number of

*Website: <http://www.rdkit.org/>

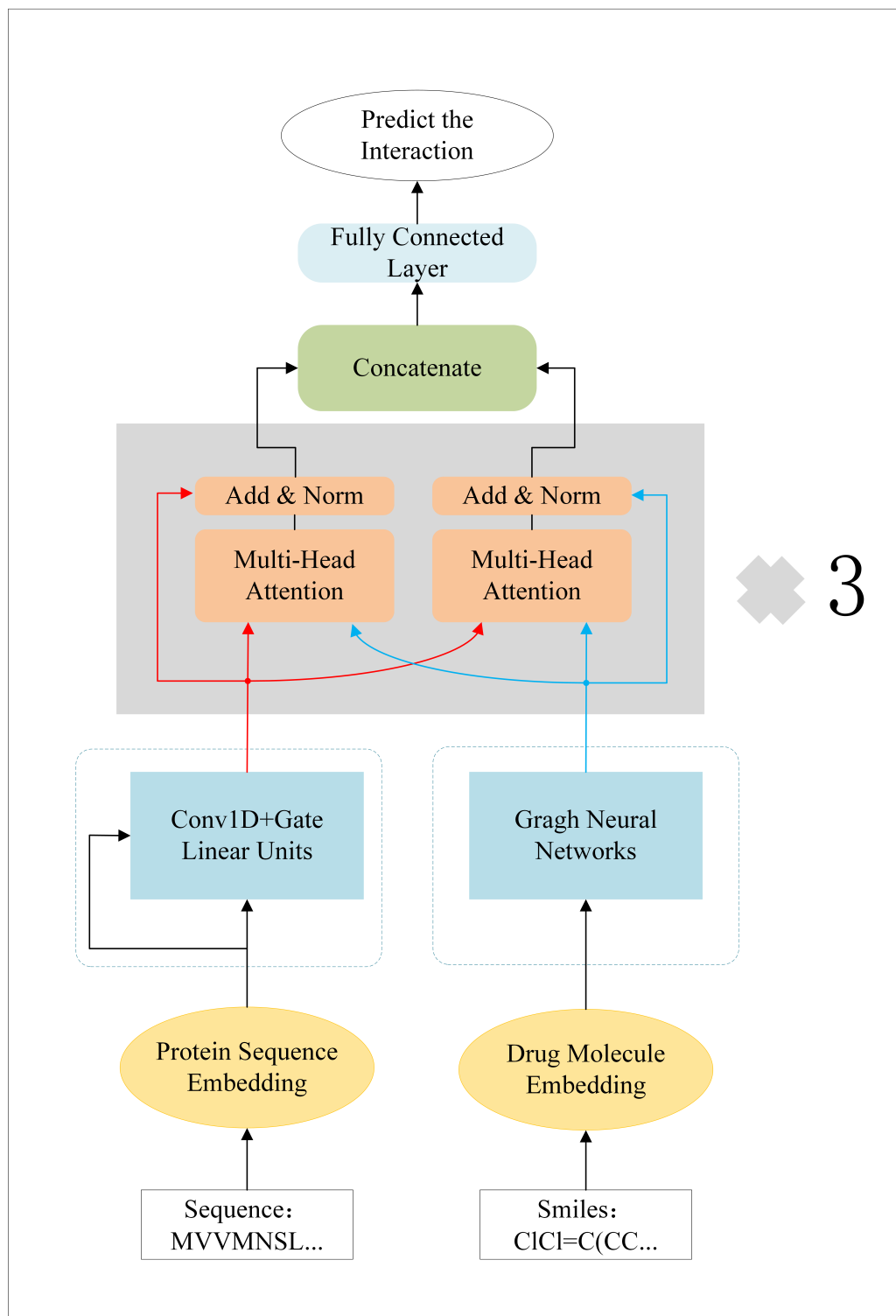


Figure 1. An overview of the entire network frame.

hops of a particular node [22, 31]. For example, we start searching from node v_i , traverse the subgraph of range r and record the information of all neighboring nodes and edges of node v_i in the subgraph. We define subgraph for node v_i in the depth r range as follows:

$$G_i^{sub^r} = (V_i^{sub^r}, E_i^{sub^r}) \quad (2.1)$$

$$E_i^{sub^r} = \{e_{mn} | m, n \in N(i, r)\} \quad (2.2)$$

where $N(i, r)$ is the set of nodes adjacent to v_i in the subgraph $G_i^{sub^r}$, including v_i . e_{mn} is the edge connecting v_m and node v_n .

According to the subgraph $G_i^{sub^r}$, we can extract the corresponding chemical features, such as atomic type, atomicity, aromaticity, etc. The details are shown in Table 1.

Table 1. Atomic characteristics and representation.

Feature	Representation
atom type	C, N, O, S, F, P, Cl, Br, B, H (onehot)
degree of atom	0, 1, 2, 3, 4, 5 (onehot)
number of hydrogens attached	0, 1, 2, 3, 4 (onehot)
implicit valence electrons	0, 1, 2, 3, 4, 5 (onehot)
aromaticity	0 or 1

We use the random initialized embedding of the extracted chemical features as the initial input to the GNN. We denote the embedding of the n -th layer network node v_i as $f_i^{(n)} \in \mathbb{R}^d$. In the GNN, we update $f_i^{(n)}$ according to the following equation:

$$f_i^{(n)} = \sigma \left(f_i^{(n-1)} + \sum_{j \in N(i, r)} h_{ij}^{(n-1)} \right) \quad (2.3)$$

where σ is the sigmoid function: $\sigma(x) = 1 / (1 + e^{-x})$ and $h_{ij}^{(n-1)}$ is the hidden vector between nodes v_i and v_j . This hidden vector can be computed by the following neural network:

$$h_{ij}^{(n)} = \varepsilon(\omega f_i^{(n)} + b) \quad (2.4)$$

where ε is the nonlinear activation function ReLU: $\varepsilon(x) = \max(0, x)$, $\omega \in \mathbb{R}^{d \times d}$ is the weight hyperparameter and $b \in \mathbb{R}^d$ is the deviation vector. After the GNN layer, we obtain the feature vector $c_1, c_2, c_3, \dots, c_l$ of a drug sequence, where l is the number of atoms in the sequence.

2.2. Protein model

A protein sequence consists of 20 amino acids. If we learn a protein sequence as a sentence, there are only 20 kinds of words that make up the sentence. To increase the diversity of features, based on the n -gram language model, we define the words in a protein sequence as n -gram amino acids [22]. For a given amino acid sequence, we split it into repeated n -gram amino acid sequences. For example, we set n to 3, and the protein sequence *MVVMNSL...TSQATP* is split into

$MVV, VVM, VMN, MNS, \dots, TSQ, SQA, QAT, ATP$, so that the variety of words composing the sentence will be expanded to 20^3 .

To ensure a reasonable vocabulary, we set n to 3. For a given protein sequence $S = a_1a_2a_3 \dots a_L$, where L is the length of the protein sequence and a_i is the i -th amino acid. We split it into:

$$[a_1a_2a_3], [a_2a_3a_4], \dots, [a_{L-2}a_{L-1}a_L]$$

We use $a_{i:i+2} \in \mathbb{R}^d$ to denote the d -dimensional embedding of the word $[a_i a_{i+1} a_{i+2}]$. We do the initialize d -dimensional embedding of the protein sequence processed by the above method and then input it into a gated convolution network with Conv1D and gated linear units [32]. We compute the hidden layer according to Eq (2.5):

$$L_i(X) = (X \times \omega_1 + s) \otimes \sigma(X \times \omega_2 + t) \quad (2.5)$$

where L_i is the i -layer in the gated convolution network, $X \in \mathbb{R}^{n \times d_1}$ is the input to the i -layer, $\omega_1 \in \mathbb{R}^{d_1 \times d_2}$, $s \in \mathbb{R}^{d_2}$, $\omega_2 \in \mathbb{R}^{d_1 \times d_2}$ and $t \in \mathbb{R}^{d_2}$ are the learning parameters, n is the length of the sequence. d_1, d_2 are the dimensions of the input and hidden features respectively, σ is sigmoid function and \otimes is matrix product. The output of the gated convolution network is the final representation of the protein sequence.

2.3. Mutual model

We extracted feature vectors of drug and protein sequences using the drug and protein modules and inputted them into the Transformer decoder. The decoder learned mutual features, resulting in drug and protein sequences with interaction features as output. Since the order of the feature vectors has no effect on the DTI modeling, we remove the positional embedding in the Transformer. The key technology of the decoder is the multi-headed self-attention layer. The multi-headed self-attention layer consists of several scaled point-attention layers for extracting the interaction information between the encoder and the decoder. The self-attention layer accepts three inputs, i.e., key K , value V and query Q , and computes the attention in the following manner:

$$Self_attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.6)$$

where d_k is a scaling factor that depends on the number of layers. Considering the complex reaction processes involving non-covalent chemical bonds within drugs and proteins, we added a module to extract interaction features using a multi-headed self-attentive layer in the Transformer decoder. The decoder takes the drug and protein sequences as inputs, enabling the extraction of drug- and protein-dominated interaction features simultaneously. The module further extracts complex interaction feature vectors between atoms and amino acids within the sequence, as depicted in Figure 1.

After the interaction features are extracted by the decoder, we obtain the interaction feature matrices $D \in \mathbb{R}^{b \times n_1 \times d}$ and $P \in \mathbb{R}^{b \times n_2 \times d}$ for the drugs and proteins. Where b is the batch size, n_1, n_2 is the number of words and d is the feature dimension. Averaging over the different dimensions of the feature matrix :

$$D_a = mean(D, 1) \quad (2.7)$$

$$P_a = mean(P, 1) \quad (2.8)$$

where $mean(input, dim)$ is a mean operation that returns the mean value of each row in the given dimension dim . The obtained feature vectors D_a and P_a are concatenated and fed to the classification block.

2.4. Classified block

The classification module consists of a multilayer fully connected neural network with an activation function of ReLU and a final layer output representing the interaction probability \hat{y} . As a binary classification task, we use binary cross-entropy loss to train Mutual-DTI:

$$Loss = - [y \log \hat{y} + (1 - y) \log (1 - \hat{y})] \quad (2.9)$$

Mutual-DTI was implemented with Pytorch 1.10.0. The original transformer model had 6 layers and contained 512 hidden dimensions, we reduced the number of layers from 6 to 3, the number of hidden layers from 512 to 10, the number of protein representations, atom representations, hidden layers and y interactions to 10 and the number of attention heads to 2, as this configuration achieves excellent generalization capabilities. During training, we used the Adam optimizer [33] with the learning rate set to 0.005 and the batch size set to 128. All settings and hyperparameters of Mutual-DTI are shown in the Table 2.

Table 2. Hyperparameters of Mutual-DTI.

Name	Value
Dimension of atom representation	10
Dimension of protein representation	10
Number of decoder layers	3
Number of hidden layers	10
Number of attention heads	2
Learning rate	5e-3
Weight decay	1e-6
Dropout	0.1
Batch size	128

3. Results

3.1. Performance on public datasets

The human dataset and the *C.elegans* dataset were created by Liu et al. [34]. These two datasets comprise of compound-protein pairs, which include both negative and positive samples that are highly plausible. The human dataset comprises 3369 positive interactions between 1052 unique compounds and 852 unique proteins, while the *C. elegans* dataset comprises 4000 positive interactions between 1434 unique compounds and 2504 unique proteins. As shown in Table 3. We randomly divided into training set, validation set and test set in the ratio of 8:1:1. In addition, we utilized AUC, precision and recall as evaluation metrics for Mutual-DTI, and compared it with some traditional machine learning methods on both the human and *C. elegans* datasets. These traditional machine learning methods for

comparison are k-NN, RF, L2-logistic (L2) and SVM. their results are from the original paper [34]. The main results are shown in Figure 2. Mutual-DTI outperforms the machine learning methods under both benchmark datasets.

Table 3. Summary of the datasets.

Dataset	Drugs	Proteins	Interactions	Positive	Negative
Human	1052	852	6728	3369	3359
C.elegans	1434	2504	7786	4000	3786
GPCR	5359	356	15343	7989	7354
Davis	68	379	25772	7320	18452

In other experiments, we compare the proposed method with recent deep learning methods used for DTI prediction. The methods are as follows: GNN-CPI [22], GNN-PT [27], TransformerCPI [28]. The main hyperparameters are set as follows:

GNN-CPI: vector dimensionality of vertices, edges and n-grams = 10, numbers of layers in gnn = 3, window size = 11, numbers of layers in cnn = 2, numbers of layers in output = 3.

GNN-PT: numbers of layers in gnn = 3, numbers of layers in output = 1, heads of attention = 2.

TransformerCPI: dimension of atom = 34, dimension of protein = 100, dimension of hidden = 64, number of hidden layers = 3, heads of attention = 8.

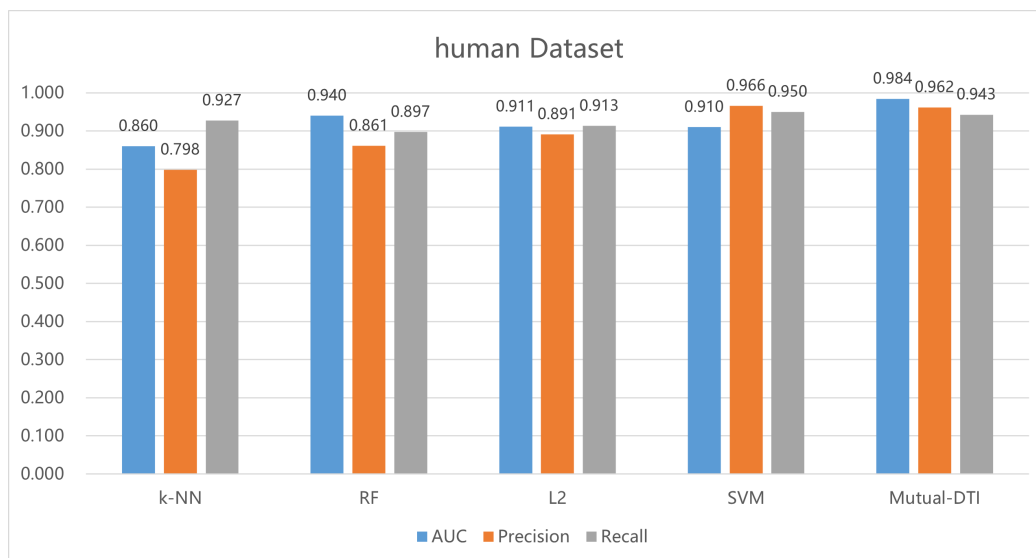
The same settings of the original paper were used for all parameter settings. We used the same pre-processing for the initial data of drugs and proteins. We preprocessed the dataset in the same way as in previous experiments, and repeated the experiment three times using different random seeds. For each repetition, we randomly split the dataset and saved the model parameters corresponding to the optimal validation set AUC for each test set. The main results under the human dataset and the C. elegans dataset are shown in Tables 4 and 5. On the human dataset, the average evaluation metrics of Mutual-DTI are 0.984, 0.962 and 0.943 for AUC, precision and recall, respectively, which outperform the other methods. On the C. elegans dataset, the average evaluation metrics AUC, precision and recall of Mutual-DTI are 0.987, 0.948 and 0.949, respectively, which mostly outperform the other models. The results suggest that Mutual-DTI can effectively learn informative features for predicting interactions from both one-dimensional protein sequences and two-dimensional molecular maps, demonstrating its generalizability across different datasets.

Table 4. Comparison between model and baseline on human dataset.

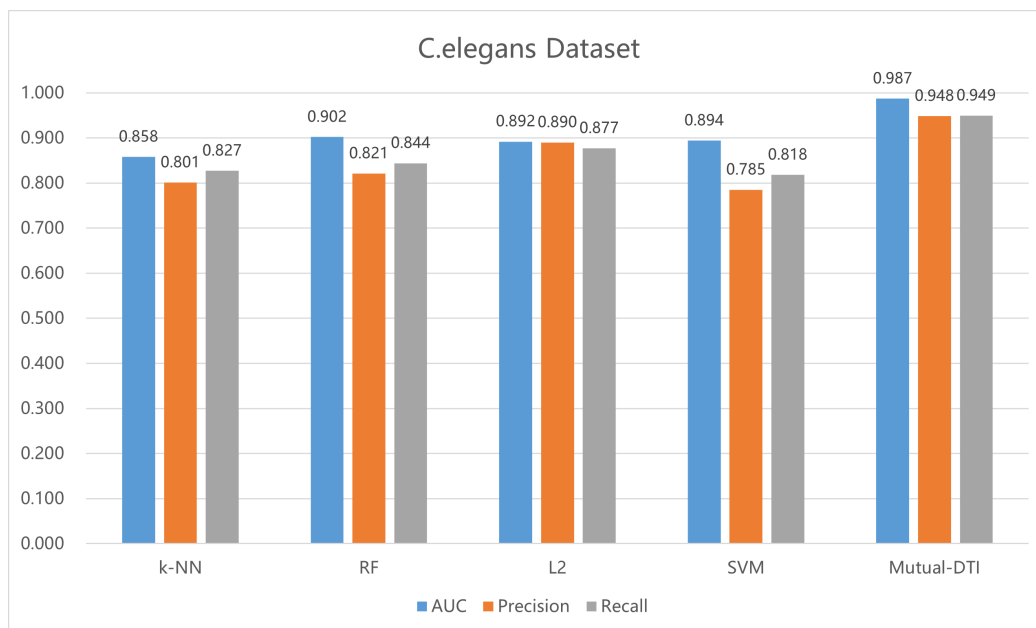
Methods	AUC	Precision	Recall
GNN-CPI	0.917 ± 0.072	0.783 ± 0.061	0.889 ± 0.096
GNN-PT	0.978 ± 0.006	0.939 ± 0.010	0.934 ± 0.006
TransformerCPI	0.972 ± 0.005	0.938 ± 0.018	0.932 ± 0.001
Mutual-DTI	0.984 ± 0.001	0.962 ± 0.019	0.943 ± 0.016

3.2. Effectiveness of mutual feature module

To evaluate the importance of interactive feature modules, we propose two sub-models. The first one is no-mutual block, which has no mutual feature module, another network with mutual feature



(a)



(b)

Figure 2. In k-NN, RF, L2, SVM and Mutual-DTI of human dataset and C.elegans dataset, AUC, accuracy and recall, where (a) is the human dataset and (b) is the C.elegans dataset.

Table 5. Comparison between model and baseline on C.elegans dataset.

Methods	AUC	Precision	Recall
GNN-CPI	0.899 ± 0.104	0.850 ± 0.132	0.778 ± 0.192
GNN-PT	0.984 ± 0.007	0.940 ± 0.024	0.933 ± 0.014
TransformerCPI	0.984 ± 0.004	0.943 ± 0.025	0.951 ± 0.016
Mutual-DTI	0.987 ± 0.004	0.948 ± 0.018	0.949 ± 0.013

module. In order to improve the accuracy of the experimental results, we evaluated a more strictly divided GPCR dataset [28]. As shown in Table 3. The key to constructing this dataset is that the drugs in its training set appear in only one class of samples (positive interaction or negative interaction DTI pairs), and in the test set appear in only the opposite class of samples. This forces the model to use protein information to learn interaction patterns and make opposite predictions for selected drugs, which is more realistic.

Table 6 shows the prediction performance of the two models on the GPCR dataset. As shown in the table, by comparing models with and without mutual features module, it can be concluded that improvements can indeed be achieved using interaction features. This suggests the need to establish correlations between drug and protein information in the DTI prediction extrapolation process. We also conducted experiments on the Davis dataset, which contains 7320 positive and 18,452 negative interactions. The Davis dataset we used was created by Zhao et al. [35]. As shown in Table 7, the model with the interaction module included also showed superior performance on the unbalanced dataset.

Table 6. Comparison between the model and no-mutual on GPCR dataset.

Methods	AUC	Precision	Recall
no-mutual-DTI	0.810 ± 0.023	0.704 ± 0.014	0.768 ± 0.030
Mutual-DTI	0.820 ± 0.014	0.699 ± 0.010	0.796 ± 0.046

Table 7. Comparison between the model and no-mutual on Davis dataset.

Methods	AUC	Precision	Recall
no-mutual-DTI	0.886 ± 0.005	0.728 ± 0.023	0.654 ± 0.005
Mutual-DTI	0.900 ± 0.002	0.767 ± 0.013	0.680 ± 0.027

3.3. Robustness of model

In this section, we employed a three-dimensional surface plot to analyze the impact of model hyperparameters (atomic and protein dimensions) on the DTI prediction performance. Since these two parameters are among the most important hyperparameters. We sampled values for atomic and protein dimensions from 10 to 40, with a gap of 10 dimensions each time. For instance, the atomic dimension/protein dimension values were: 10/10, 10/20, 10/30, ..., 20/10, 20/20, ..., 40/30, 40/40, for a total of 16 different settings, and experiments were conducted multiple times with different random seeds. Other settings were the same as the previous experiments. As shown in the Figure 3, the x-axis represents the atomic dimension, the y-axis represents the protein dimension and the z-axis represents the AUC obtained from the test set. From the results, it can be seen that under different dimension settings, the surfaces are very smooth and the model exhibits good robustness.

3.4. Model interpretation

To demonstrate that the mutual feature module we introduced not only enhances the performance of the model but also provides deeper interpretation, we conducted a case study. First, we applied a Frobenius parametric solution to the protein feature vector matrix obtained by the Transformer decoder. Next, we used the Softmax function to derive attention weights of the protein sequences, which were

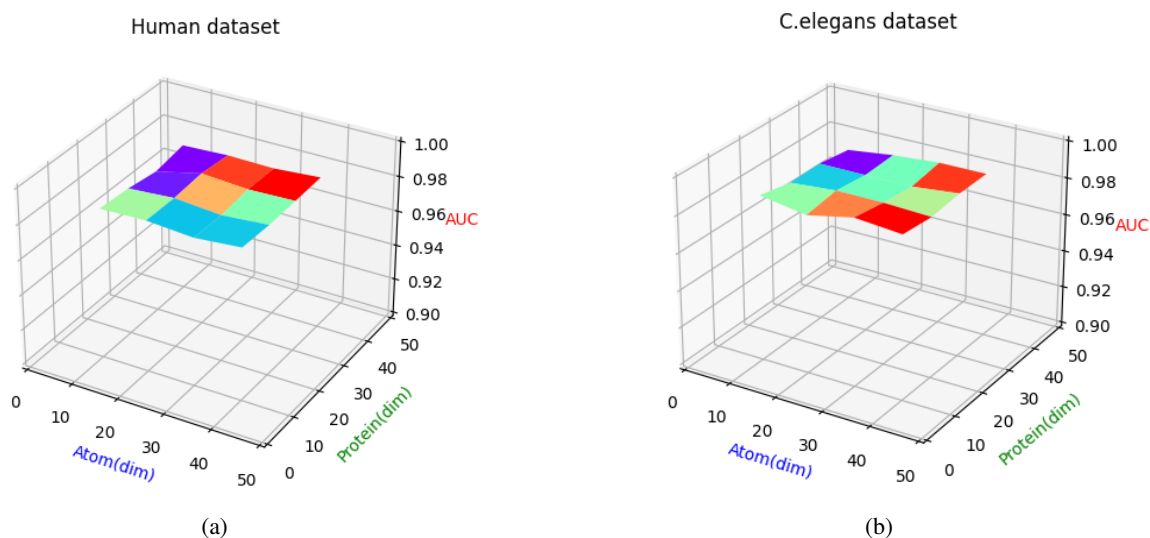


Figure 3. 3D surface plots of different dimensions on two datasets, the x-axis represents the atomic dimension, the y-axis represents the protein dimension and the z-axis represents the AUC obtained from the test set. (a) is the human dataset and (b) is the C.elegans dataset.

then mapped onto the 3D structure of the complex to visualize the regions that are more efficient for drug-protein reaction. The attention weights of the crystal structure of gw0385-bound HIV protease D545701 (PDB:2FDD) are shown in Figure 4. The complex has a total of 12 binding sites. We have marked the regions that received high attention more than 0.75 in red, and it can be found that a total of 4 of the 12 binding sites received high attention scores, namely ASP-25, ALA-28, PRO-81 and ALA82. The results show that Mutual-DTI helps to narrow down the search space of binding sites.

4. Discussion

Based on our prior experiments on robustness, it is evident that the model's AUC slightly varies when encoding and embedding atoms and proteins in various dimensions. Our inference is that the GNN module, which learns the drug molecule characteristics, and the gated convolutional unit module, which learns the protein features, can effectively extract feature information in the Mutual-DTI model. This indicates that constructing drug and protein sequences into subgraphs via local breadth-first search algorithms and constructing words via n-gram methods is reasonable. In the ablation experiment, we discovered that removing the module that extracts mutual reaction features resulted in a significant decrease in the model's prediction accuracy. Our speculation is that the model only learns each sequence's individual feature information after learning drug and protein features, while DTI is a dynamic process. Introducing the Mutual learning module treats drug and protein features as the main body, and they dynamically focus on each other's key parts in the learning layer, thus directly capturing the interaction features of the two given sequences. By learning the interaction features, the model obtains a deeper understanding of the DTI process and can more easily capture the crucial parts that may contribute to the reaction when dealing with unknown drug and protein data, leading to superior performance in predicting results.

From the perspective of model complexity analysis, Mutual-DTI has higher complexity compared to

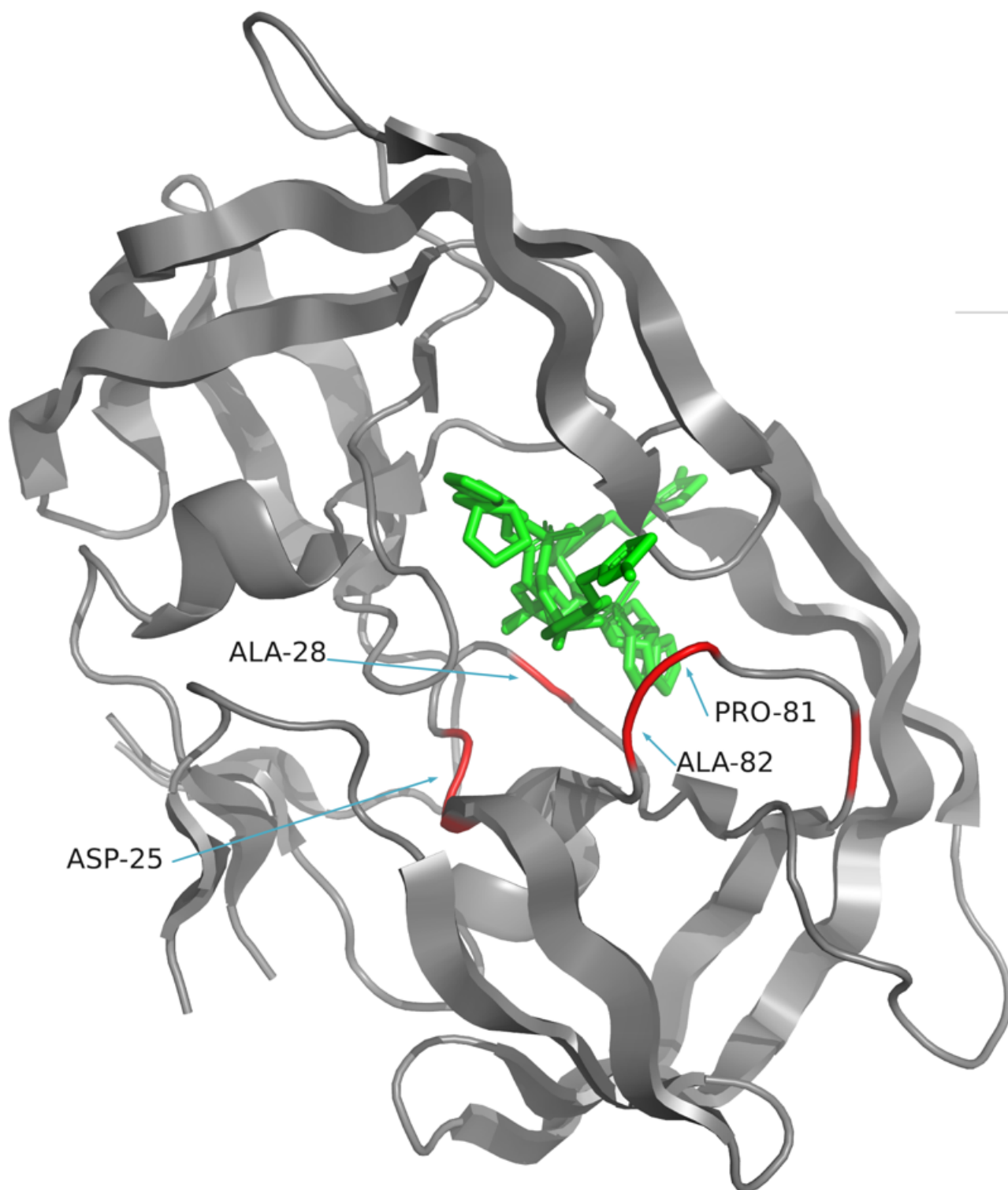


Figure 4. Attention weight of Protease (PDB: 2FDD).

the networks that only use GNN(e.g., GNN-CPI). This is because we use a more complex self-attention mechanism, which allows us to capture long-range dependencies between tokens in the sequence. Compared to TransformerCPI, which is also based on Transformer, Mutual-DTI has lower complexity. The rationale behind this is that, while Mutual-DTI takes into account two parallel multi-head attention layers, the network's number of attention heads is diminished from 8 to 2, and we devised a lower hidden layer dimension. Consequently, this has notably curtailed the number of parameters in the Mutual-DTI. These designs help Mutual-DTI better fit the training data while avoiding overfitting due to excessive complexity.

5. Conclusions

In this paper, we present a Transformer-based network model for predicting DTI and introduce a module for extracting sequence interaction features to model complex reaction processes between atoms and amino acids. To validate the effectiveness of Mutual-DTI, we compare it with the latest baseline on two benchmark datasets. The results show that Mutual-DTI outperforms the baseline. We also evaluate Mutual-DTI on the label reversal dataset and observe a significant improvement with the introduction of the mutual feature module. Finally, we map the attention weights obtained by the mutual feature module to the protein sequences, which helps us better interpret the model and determine the reliability of the predictions.

Although Mutual-DTI shows effective performance in predicting DTI, there is still room for improvement. The experimental results show a significant decrease in performance on the strictly limited label inversion dataset compared to the human dataset and the *C.elegans* dataset. This suggests that the feature extraction of sequences is very limited, and adding a 3D representation of molecules or proteins may help extract more information.

Acknowledgments

The work was supported by Open Funding Project of the State Key Laboratory of Biocatalysis and Enzyme Engineering under grant No. SKLBEE2021020 and SKLBEE2020020, the High-level Talents Fund of Hubei University of Technology under grant No. GCRC2020016, Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province under grant No. 2020E10010-02 and Natural Science Foundation of Hubei Province under grant No. 2021CFB282.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. C. Chen, H. Shi, Y. Han, Z. Jiang, X. Cui, B. Yu, DNN-DTIs: Improved drug-target interactions prediction using XGBoost feature selection and deep neural network, *Comput. Biol. Med.*, **136** (2021), 104676. <https://doi.org/10.1101/2020.08.11.247437>

2. X. Ru, X. Ye, T. Sakurai, Q. Zou, C. Lin, Current status and future prospects of drug–target interaction prediction, *Briefings Funct. Genomics*, **20** (2021), 312–322. <https://doi.org/10.1093/bfpg/elab031>
3. G. Huang, F. Yan, D. Tan, A review of computational methods for predicting drug targets, *Curr. Protein Pept. Sci.*, **19** (2018), 562–572. <https://doi.org/10.1016/j.jad.2018.12.111>
4. J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, et al., Applications of machine learning in drug discovery and development, *Nat. Rev. Drug Discovery*, **18** (2019), 463–477. <https://doi.org/10.1038/s41573-019-0024-5>
5. E. Maia, L. C. Assis, T. Oliveira, A.M. de Silva, A. G. Taranto, Structure-based virtual screening: from classical to artificial intelligence, *Front. Chem.*, **8** (2020), 343. <https://doi.org/10.3389/fchem.2020.00343>
6. H. Mubarak, S. Naomie, A. D. Mohammed, S. Faisal, A. Ali, Adapting document similarity measures for ligand-based virtual screening, *Molecules*, **21** (2016), 476. <https://doi.org/10.3390/molecules21040476>
7. R. Ferdousi, R. Safdari, Y. Omid, Computational prediction of drug–drug interactions based on drugs functional similarities, *J. Biomed. Inf.*, **70** (2017), 54. <https://doi.org/10.1016/j.jbi.2017.04.021>
8. M. Bredel, E. Jacoby, Chemogenomics: An emerging strategy for rapid target and drug discovery, *Nat. Rev. Genet.*, **5** (2004), 262–275. <https://doi.org/10.1038/nrg1317>
9. K. Bleakley, Y. Yamanishi, Supervised prediction of drug–target interactions using bipartite local models, *Bioinformatics*, **25** (2009), 2397–2403. <https://doi.org/10.1093/bioinformatics/btp433>
10. F. Cheng, Y. Zhou, J. Li, W. Li, G. Liu, Y. Tang, Prediction of chemical–protein interactions: multitarget-QSAR versus computational chemogenomic methods, *Mol. BioSyst.*, **8** (2012), 2373–2384. <https://doi.org/10.1039/C2MB25110H>
11. M. Gönen, Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorization, *Bioinformatics*, **28** (2012), 2304–2310. <https://doi.org/10.1093/bioinformatics/bts360>
12. S. Liu, J. An, J. Zhao, S. Zhao, H. Lv, S. Wang, et al., Drug-Target interaction prediction based on multisource information weighted fusion, *Contrast Media Mol. Imaging*, **2021** (2021). <https://doi.org/10.1155/2021/6044256>
13. J. Li, X. Yang, Y. Guan, Z. Pan, Prediction of drug–target interaction using dual-network integrated logistic matrix factorization and knowledge graph embedding, *Molecules*, **27** (2022), 5131. <https://doi.org/10.3390/molecules27165131>
14. Y. Ding, J. Tang, F. Guo, Q. Zou, Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization, *Brief. Bioinf.*, **23** (2022). <https://doi.org/10.1093/bib/bbab582>
15. L. Jacob, J. P. Vert, Protein-ligand interaction prediction: An improved chemogenomics approach, *Bioinformatics*, **24** (2008), 2149–2156. <https://doi.org/10.1093/bioinformatics/btn409>

16. T. Van Laarhoven, S. B. Nabuurs, E. Marchiori, Gaussian interaction profile kernels for predicting drug–target interaction, *Bioinformatics*, **27** (2011), 3036–3043. <https://doi.org/10.1093/bioinformatics/btr500>
17. F. Wang, D. Liu, H. Wang, C. Luo, M. Zheng, H. Liu, et al., Computational screening for active compounds targeting protein sequences: Methodology and experimental validation, *J. Chem. Inf. Model.*, **51** (2011), 2821–2828. <https://doi.org/10.1021/ci200264h>
18. Y. Wang, J. Zeng, Predicting drug–target interactions using restricted Boltzmann machines, *Bioinformatics*, **29** (2013), i126–i134. <https://doi.org/10.1093/bioinformatics/btt234>
19. Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, *Bioinformatics*, **24** (2008), i232–i240. <https://doi.org/10.1093/bioinformatics/btn162>
20. K. Tian, M. Shao, Y. Wang, J. Guan, S. Zhou, Boosting compound–protein interaction prediction by deep learning, *Methods*, **110** (2016), 64–72. <https://doi.org/10.1016/j.ymeth.2016.06.024>
21. I. Lee, J. Keum, H. Nam, DeepConv-DTI: Prediction of drug–target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.*, **15** (2019), e1007129. <https://doi.org/10.1371/journal.pcbi.1007129>
22. M. Tsubaki, K. Tomii, J. Sese, Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, *Bioinformatics*, **35** (2019), 309–318. <https://doi.org/10.1093/bioinformatics/bty535>
23. S. Li, F. Wan, H. Shu, T. Jiang, D. Zhao, J. Zeng, MONN: A multi-objective neural network for predicting compound–protein interactions and affinities, *Cell Syst.*, **10** (2020), 308–322. <https://doi.org/10.1016/j.cels.2020.03.002>
24. R. Zamora-Resendiz, S. Crivelli, Structural learning of proteins using graph convolutional neural networks, *BioRxiv*, (2019), 610444. <https://doi.org/10.1101/610444>
25. S. Ryu, J. Lim, S. H. Hong, W. Y. Kim, Deeply learning molecular structure–property relationships using attention-and gate-augmented graph convolutional network, preprint, arXiv: 1805.10988. <https://doi.org/10.48550/arXiv.1805.10988>
26. X. Ru, X. Ye, T. Sakurai, Q. Zou, NerLTR-DTA: Drug–target binding affinity prediction based on neighbor relationship and learning to rank, *Bioinformatics*, **38** (2022), 1964–1971. <https://doi.org/10.1093/bioinformatics/btac048>
27. J. Wang, X. Li, H. Zhang, GNN-PT: Enhanced prediction of compound–protein interactions by integrating protein transformer, preprint, arXiv: 2009.00805. <https://doi.org/10.48550/arXiv.2009.00805>
28. L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, et al., TransformerCPI: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments, *Bioinformatics*, **36** (2020), 4406–4414. <https://doi.org/10.1093/bioinformatics/btaa524>

29. Z. H. Ren, Z. H. You, Q. Zou, C. Q. Yu, Y. F. Ma, Y. J. Guan, et al., DeepMPF: Deep learning framework for predicting drug–target interactions based on multi-modal representation with meta-path semantic analysis, *J. Transl. Med.*, **21** (2023), 1–18. <https://doi.org/10.1186/s12967-023-03876-3>
30. Y. Wu, L. Zhu, Y. Yan, Y. Yang, Dual attention matching for audio-visual event localization, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 6292–6300.
31. F. Costa, K. De Grave, Fast neighborhood subgraph pairwise distance kernel, in *Proceedings of the 26th International Conference on Machine Learning*, (2010), 255–262.
32. Y. N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks, in *International Conference on Machine Learning*, (2017), 933–941.
33. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, arXiv: 1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>
34. H. Liu, J. Sun, J. Guan, J. Zheng, S. Zhou, Improving compound–protein interaction prediction by building up highly credible negative samples, *Bioinformatics*, **31** (2015), i221–i229. <https://doi.org/10.1093/bioinformatics/btv256>
35. Q. Zhao, H. Zhao, K. Zheng, J. Wang, HyperAttentionDTI: Improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism, *Bioinformatics*, **38** (2022), 655–662. <https://doi.org/10.1093/bioinformatics/btab715>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)