Bioinformatics, 38(17), 2022, 4153–4161
https://doi.org/10.1093/bioinformatics/btac485
Advance Access Publication Date: 8 July 2022
Original Paper

Systems biology

IIFDTI: predicting drug–target interactions through interactive and independent features based on attention mechanism

Zhongjian Cheng[1], Qichang Zhao [1], Yaohang Li[2] and Jianxin Wang [1,*]

[1]Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China and [2]Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA

*To whom correspondence should be addressed.

Associate Editor: Karsten Borgwardt

Received on December 7, 2021; revised on May 2, 2022; editorial decision on June 20, 2022; accepted on July 7, 2022

Abstract

Motivation: Identifying drug–target interactions is a crucial step for drug discovery and design. Traditional biochemical experiments are credible to accurately validate drug–target interactions. However, they are also extremely laborious, time-consuming and expensive. With the collection of more validated biomedical data and the advancement of computing technology, the computational methods based on chemogenomics gradually attract more attention, which guide the experimental verifications.

Results: In this study, we propose an end-to-end deep learning-based method named IIFDTI to predict drug–target interactions (DTIs) based on independent features of drug–target pairs and interactive features of their substructures. First, the interactive features of substructures between drugs and targets are extracted by the bidirectional encoder–decoder architecture. The independent features of drugs and targets are extracted by the graph neural networks and convolutional neural networks, respectively. Then, all extracted features are fused and inputted into fully connected dense layers in downstream tasks for predicting DTIs. IIFDTI takes into account the independent features of drugs/targets and simulates the interactive features of the substructures from the biological perspective. Multiple experiments show that IIFDTI outperforms the state-of-the-art methods in terms of the area under the receiver operating characteristics curve (AUC), the area under the precision-recall curve (AUPR), precision, and recall on benchmark datasets. In addition, the mapped visualizations of attention weights indicate that IIFDTI has learned the biological knowledge insights, and two case studies illustrate the capabilities of IIFDTI in practical applications.

Availability and implementation: The data and codes underlying this article are available in Github at https://github.com/czjczj/IIFDTI.

Contact: jxwang@mail.csu.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Identifying potential drug–target interactions (DTIs) plays an important role in drug discovery and drug reposition (Ezzat *et al.*, 2019). Although traditional biological experiments are trustworthy, they are typically costly and laborious. With massive biomedical data and knowledge being applicable, these methods based on chemogenomic begin to receive attention from researchers. Chemogenomic-based approaches use information from both the drug and target sides simultaneously to perform predictions. Therefore, many machine learning-based methods have been proposed since then (Bleakley and Yamanishi, 2009; Cheng *et al.*, 2012; Mohamed *et al.*, 2020; Olayan *et al.*, 2018; Wang and Zeng, 2013; Wang *et al.*, 2011; Yuan *et al.*, 2016; Zong *et al.*, 2017). Combining the advantages of the feature-based and similarity-based methods, Yuan *et al.* (2016) propose an ensemble-based method (called DrugE-Rank) to improve the prediction performance. Based on DTI heterogeneous graph, Olayan *et al.* (2018) develop a machine learning-based method (named DDR) which applies the non-linear similarity fusion method to combine different types of similarities. Formulating the problems as link predictions in knowledge graphs, Mohamed *et al.* (2020) propose a model called TriModel based on the knowledge graph embedding. Then, the embedding vectors are consequently used to infer candidate DTIs based on their scores.

With the rapid development of deep learning, many end-to-end deep learning-based methods have been applied for predicting DTIs (Chen *et al.*, 2020; Gao *et al.*, 2018; Huang *et al.*, 2021; Karimi

et al., 2019; Lee et al., 2019; Nguyen et al., 2021; Öztürk et al., 2018; Rifaioglu et al., 2021; Tsubaki et al., 2019; Wan et al., 2019; Zheng et al., 2020). According to the different ways of extracting features of drugs and targets, these methods can be divided into two types: one is based on independent features of drug–target pairs, and the other is based on the interaction information of local substructures. As shown in Figure 1, methods based on independent features regard drugs and targets as two independent parts. They extract the feature vectors of drugs and targets using independent feature extractors. Then, the feature vectors are fused and inputted into the downstream tasks. Gao et al. (2018) propose an interpretable architecture to predict DTIs from low-level representations, in which the characteristics of drugs and targets are extracted by long short-term memory (LSTM) and graph convolutional network, respectively. Öztürk et al. (2018) propose a model named DeepDTA which uses only sequence information to predict drug–target binding affinities. DeepDTA models both the characterizations of drugs and targets with convolutional neural network (CNN). By using a variety of graph-based models, Nguyen et al. (2021) propose the model named GraphDTA to improve the predictive ability of DTA by capturing the structural information of drugs. Following the visual question answering mode, Zheng et al. (2020) propose the model named drugVQA to predict DTIs. DrugVQA separately extracts the features of variable-length distance maps and linear notations by CNN and self-attention Bi-LSTM, respectively.

Although methods based on independent features of drugs and targets have achieved good performance, the drug–target interactions include the interactions of local substructures of the drug–target pairs in high-dimensional spaces (Schenone et al., 2013). Therefore, it is meaningful to combine interactive information about drug–target pairs for predictions. Methods based on the interactive information of local substructures take into account of the real biological DTI processes and simulate them. During the processes of feature extractions, drugs and targets are no longer considered as independent parts. These methods can model the interactive features of local substructures of drug–target pairs as input for the downstream tasks. Chen et al. (2020) propose a novel transformer-based neural network named TransformerCPI. TranformerCPI reduces some specific pitfalls for sequence-based DTI models. By developing knowledge inspired sub-structural pattern mining algorithm and interaction modeling module, Huang et al. (2021) propose an interpretable model called MolTrans. MolTrans is capable of explaining the prediction results from reasonable cues.

Although it is reasonable to focus on the interactive information between drugs and targets, drug–target interaction is a complex process involving the knowledge of biology and chemistry. Extracting independent features for drug–target pairs may also provide more discriminative information for predictions. We cannot simply separate independent features from interactive features, but systematically extract effective features from them. To address the above issues, in this study, we propose a novel deep learning-based model named IIFDTI. It fuses the interactive and independent features between drugs and targets to predict DTIs. More specifically, we simulate the actual hypothesis from biological inspirations and extract the interactive features of local substructures for drug–target pairs. For this process, the bidirectional encoder–decoder structure is chosen as the extractor, which has rich semantic extraction capabilities. Simultaneously, we model the independent features of drugs and targets, which are fused with interactive features for classifications downstream. Among them, the graph attention network (GAT) is used to extract the independent features of drugs, and it can capture the topological information between atoms.

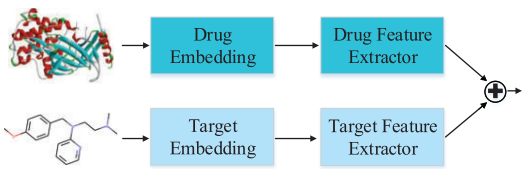

Fig. 1. An overview of the methods based on only independent features

Similarly, the convolutional structures with different kernel scales are used to extract independent features of targets. Finally, four feature vectors are fused and inputted into fully connected dense layers for predicting DTIs. Multiple experimental results show that IIFDTI outperforms four state-of-the-art methods on five different datasets in terms of AUC, AUPR, precision and recall. To summarize, the main contributions of our works are listed as follows:

- We propose a novel model named IIFDTI for predicting DTIs. To our knowledge, it is the first to consider interactive features of local substructures of drug–target pairs and independent features for both drugs and targets.
- We design rich ablation experiments to verify the significance of the combination of independent feature extraction and interactive feature extraction for DTIs prediction.
- Our comprehensive experimental results demonstrate the proposed method outperforms four state-of-the-art methods on five benchmark datasets. Simultaneously, the visualizations of attention weights and case studies indicate that the proposed model can provide biological insights.

2 Materials and methods

2.1 Benchmark datasets

We mainly construct the dataset of experiments from DrugBank database(version 5.1.5) (Wishart et al., 2006). After processing the dataset, the DrugBank dataset which contains 6645 drugs, 4254 targets and 17 511 known drug–protein interactions are obtained. Following some common practice (Huang et al., 2021), these negative samples of the DrugBank dataset are randomly sampled from unknown drug–protein pairs and kept the same number as positive samples. Moreover, we also apply our model on some previous benchmark datasets including Human dataset (Liu et al., 2015), *Caenorhabditis elegans* dataset (Liu et al., 2015), BindingDB dataset (Gao et al., 2018), GPCR dataset (Chen et al., 2020). Different from randomly selecting negative samples, Human and *C.elegans* datasets are constructed based on the highly credible negative DTI samples using a systematic screening framework (Liu et al., 2015). As shown in Table 1, the Human dataset contains 3369 positive interactions between 1052 drugs and 852 targets; *C.elegans* dataset contains 4000 positive interactions between 1434 drugs and 2504 targets. BindingDB dataset is constructed from (Gilson et al., 2016), and the training, validation and test sets of it are well-designed. It contains 33 772 positive interactions and 27 486 negative interactions, whose detailed dataset division is listed in Supplementary Section S1. GPCR dataset is constructed from GLASS database (Chan et al., 2015), which is originally used for predicting drug–target affinity (DTA). Divided into positive samples and negative samples by thresholds of 6.0, GPCR dataset contains 15 343 interactions between 5359 drugs and 356 targets. The training and test sets of GPCR dataset are well-designed, and the detailed dataset information is listed in Supplementary Section S1.

2.2 Methods

In this study, we propose a novel approach named IIFDTI to predict DTIs. We first give a brief description of the approach, and then focus on the different modules of the approach.

Table 1. Summary of the benchmark datasets

Datasets	Targets	Drugs	Interactions	Positive	Negative
DrugBank	4254	6645	35 022	17 511	17 511
Human	852	1052	6738	3369	3369
C.elegans	2504	1434	8000	4000	4000
BingDingDB	812	49 745	61 258	33 772	27 486
GPCR	356	5359	15 343	7989	7354

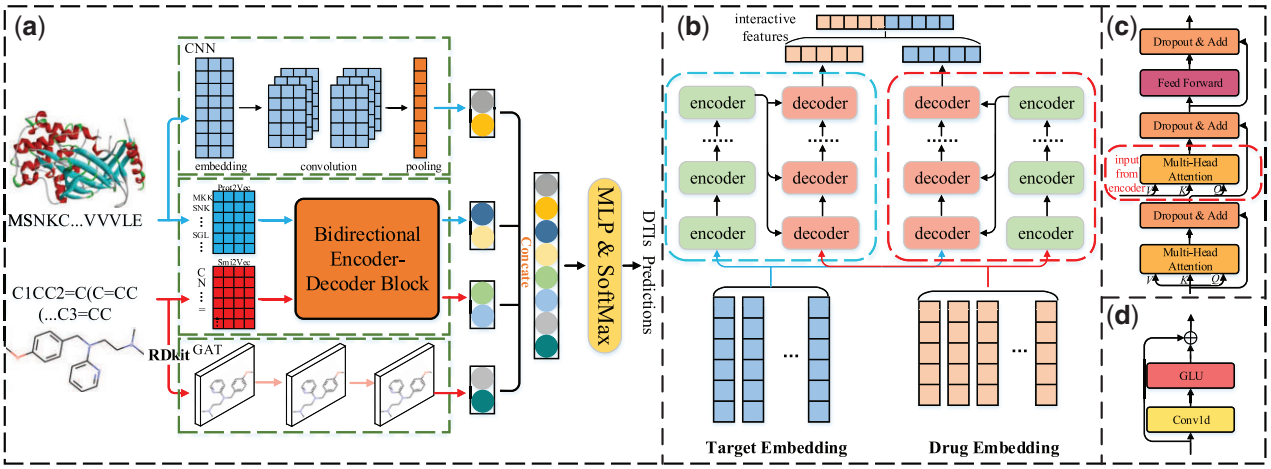

Fig. 2. The network architecture of IIFDTI. (**a**) The overview of the proposed prediction approach. (**b**) The detailed bidirectional encoder–decoder block structure in (a). (**c**) The network structures of the decoder in (b). (**d**) The network structures of the encoder in (b)

Figure 2 shows the network architecture of IIFDTI. It is mainly composed of independent feature extraction module, interactive feature extraction module and prediction module. For the independent feature extraction module, the GAT and multi-scale one-dimensional convolution are used to capture the independent features of drugs and targets, respectively. For the interactive feature extraction module, the bidirectional encoder–decoder is designed to extract the interactive features of drug–target pairs. Finally, the extracted independent and interactive features are inputted into the prediction module to obtain the prediction results.

2.2.1 Interactive features of local substructures of drug–target pairs

With the rapid development of natural language processing, many powerful models and neural network structures have been well proposed, such as BERT (Devlin *et al.*, 2019), Transformer-XL (Dai *et al.*, 2019) and Longformer (Beltagy *et al.*, 2020). They are all designed and improved based on transformer (Vaswani *et al.*, 2017), which has an encoder–decoder-based structure. In the decoder of transformer structure, the interactive features of two sequences are captured. It is natural to extract the interactive features of drug–target pairs. (Chen *et al.*, 2020) proposes a unidirectional interactive feature-based model for predicting drug–target interactions based on transformer structure. However, the inputs of drug–target interaction tasks contain two different substances (e.g. drugs and targets), the unidirectional interactive feature is insufficient for feature extractions. Inspired by this, we construct the bidirectional encoder–decoder structure to extract interactive features, which is shown in Figure 2b and takes drugs and targets as inputs.

For the raw data of amino acid sequences and SMILES strings, the sequences are embedded into real-valued vectors by word2vec (Mikolov *et al.*, 2013), which preserves the semantic relationships among symbol levels. For an amino acid sequence, it can split into overlapping 3-g sequences. Similarly, each character in the SMILES string is split to represent as a word. From the pre-trained embedding dictionaries [Smi2Vec (Quan *et al.*, 2018) and Prot2Vec (Asgari and Mofrad, 2015)], each amino acid sequence and each SMILES string can be transformed to the matrix $P \in R^{l_p \times h_p}$ and $D \in R^{l_d \times h_d}$, respectively. l_p and l_d are the length of target sequences and drug sequences, respectively. h_p and h_d are the dimensions of feature vectors for targets and drugs, respectively. The bidirectional encoder–decoder block takes P and D as inputs and it contains two directions (i.e. drug and target direction). For the target direction, the drug feature matrix D and the target feature matrix P are inputted to the encoder and decoder, respectively. For the drug direction, D and P are inputted to the decoder and encoder, respectively.

The structure of the encoder is constituted by the multi-layer combinations of one-dimensional convolutions and gated linear units (Dauphin *et al.*, 2017). It ensures that effective features can be extracted while preventing the model from overfitting due to complex structures. Then, we can get the feature matrices of drugs and targets after the ith encoder of the encoder–decoder framework, which is described as follows:

$$D_{i+1}^{encoder} = \delta(D_i^{encoder} * W1_i^{drug} + b1_i^{drug}) \odot (D_i^{encoder} * W2_i^{drug} + b2_i^{drug}) \quad (1)$$

$$P_{i+1}^{encoder} = \delta(P_i^{encoder} * W1_i^{prot} + b1_i^{prot}) \odot (P_i^{encoder} * W2_i^{prot} + b2_i^{prot}) \quad (2)$$

where $\delta()$ is the sigmoid function. $*$ is the one-dimensional convolution operation. $\odot$ is the element-wise product operation. $W1_i^{drug}, W2_i^{drug} \in R^{k_d \times m \times m}$, $b1_i^{drug}$, $b2_i^{drug} \in R^m$, $W1_i^{prot}, W2_i^{prot} \in R^{k_p \times m \times m}$ and $b1_i^{prot}, b2_i^{prot} \in R^m$ are the learnable parameters. k_p and k_d are the filter sizes in encoders for target sequences and drug sequences, respectively. m is the feature dimension of hidden layers. $D_i^{encoder} \in R^{l_d \times m}$ and $P_i^{encoder} \in R^{l_p \times m}$ are the inputs to the ith encoder. In particular, $D_1^{encoder} = D \cdot W_e^{drug}$ and $P_1^{encoder} = P \cdot W_e^{prot}$, and $W_e^{drug} \in R^{h_d \times m}$ and $W_e^{prot} \in R^{h_p \times m}$ are the learnable parameters. $D_{i+1}^{encoder}$ and $P_{i+1}^{encoder}$ are the outputs of the ith encoder in the encoder–decoder framework, where $i \in \{1, \ldots, L\}$ and L is the number of encoder layers. Finally, we can get feature matrices of substructures for drugs and targets ($D_{L+1}^{encoder}$, $P_{L+1}^{encoder}$) after convolutions by multiple encoders.

As shown in Figure 2c, decoder is mainly composed of multi-head attention, dropout, residual connection (He *et al.*, 2016), and position-wise feed-forward networks. The important operation of multi-head attention is the function $Attention(Q, K, V)$, whose inputs are three feature matrices ($Q \in R^{l_q \times d_k}$, $K \in R^{l_k \times d_k}$, and $V \in R^{l_v \times d_v}$), where l_q, l_k and l_v are the dimension of the input length. d_k and d_v indicate the transformed dimensions. Then, we can get the output matrix as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (3)$$

The multi-head attention contains h heads, where the ith head is computed as follows:

$$head_i = \text{Attention}(Q_i, K_i, V_i). \quad (4)$$

The multi-head attention concatenates the results from h heads as follows:

$$\text{MultiHead}(Q_{1\ldots h}, K_{1\ldots h}, V_{1\ldots h}) = \text{Concat}(head_1, \ldots, head_h)W^O \quad (5)$$

where $W^O \in R^{(h \cdot d_v) \times m}$ are learnable parameter matrices. h is the

number of heads. The multi-head attention can make the model more robust by combining information from different representation subspaces. As shown in Figure 2c, the decoder contains two multi-head attention modules. For the bottom one, the Q, K and V are the mappings of input of the decoder. For the top one, the K and V are the output from the encoder, and the Q is from the processing stream of decoder. As shown in the red-dashed box in Figure 2c, through the calculation process for attention weights in multi-head attention, the decoder can dynamically capture the interactive information between the outputs of the encoder and input of the decoder. The detailed example for extracting interactive features between substructures of the drug–target pair is listed in Supplementary Section S2, and the detailed inputs and outputs of the bidirectional encoder–decoder block on both directions are listed in Supplementary Section S3.

For the target direction, we get the feature matrix $\{p_1, p_2, \ldots\ldots, p_{l_p}\}$ and the target-side interactive feature L_{prot} is computed as follows:

$$L_{prot} = \sum_{i=1}^{l_p} p_i \tag{6}$$

Similarly, for the drug direction, we get the feature matrix $\{d_1, d_2, \ldots\ldots, d_{l_d}\}$ and the drug-side interactive feature L_{drug} is computed as follows:

$$L_{drug} = \sum_{i=1}^{l_d} d_i \tag{7}$$

2.2.2 Independent features of drugs and targets

2.2.2.1 Representation learning for drugs. In this section, the topological information in the drug molecular graph is extracted as independent features. It is complementary to the sequence representation of drugs in Section 2.2.1. For each molecule structure, it is represented as a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ from SMILES string by RDKit tools. In the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, $\mathcal{V}$ is the set of all atoms in a molecule, and each atom is represented as the f-dimensional feature vector $x_i \in R^f, \forall i \in \mathcal{V}$, which includes the physicochemical properties. The detailed features are listed in Supplementary Section S4. f is 34 in the implementation. $\mathcal{E}$ is the set of covalent chemical bonds. Then, the topological structure information of drug molecules is extracted by T-layer GAT (Velickovic *et al.*, 2018), where T is the total number of layers for GAT. The detailed pseudo-codes of applying GAT on drug molecules are listed in Section S5. During the tth convolution, the ith atom of molecule graphs calculates the attention weight a_{ij}^t with its neighbor nodes $j \in \mathcal{N}(i)$. Then the ith atom vector can be updated by gathering neighbor information based on attention weights. Finally, each atom contains the T-radius subgraph topological information after T convolutions. The aggregate representation of a drug molecule is substituted as G_{drug}.

$$G_{drug} = \sum_i Z_i, \quad i \in \mathcal{V} \tag{8}$$

where Z_i is the feature vector of the ith vertex in the graph after T convolutions.

2.2.2.2 Representation learning for targets. To extract the independent information of targets, the multi-scale one-dimensional convolution is used to capture the features of targets. Different from the embedding in the bidirectional encoder–decoder block, a corpus containing all human protein sequences in Uniprot database is used to train word2vec model (Chen *et al.*, 2020; Mikolov *et al.*, 2013). Finally, the pre-trained word list with the word vector dimension of 100 is obtained to initialize the CNN embedding. Then, we can get the feature matrix of the target sequence $P_{cnn} \in R^{l_p \times d_p}$ after embedding, where d_p is the dimension of the embedding vector. Multiple filters with different convolution kernels are used to extract the representative features of semantic information, and the width of the sliding window for the ith convolution kernel is κ_i.

$$C_i = \sigma(P_{cnn} * W_i^{cnn} + b_i^{cnn}) \tag{9}$$

where $W_i^{cnn} \in R^{\kappa_i \times d_p \times m}, i \in \{1, \ldots, n\}$, $b_i^{cnn} \in R^m$ are the learnable parameters, and m is the feature dimension of hidden layers. n is the number of different convolution kernels. $\sigma()$ is the non-linear function. Furthermore, the most important feature $\widetilde{c_i} = \max\{C_i\}$ is captured by max-pooling. During the training process, the BatchNorm (Ioffe and Szegedy, 2015) is used to make the model converge efficiently. Finally, we can get the independent features of the target G_{prot}.

$$G_{prot} = MLP\left(Concat(\widetilde{c_1}, \ldots, \widetilde{c_n})\right) \tag{10}$$

where $MLP()$ is the multi-layer nonlinear transformation function (Huang *et al.*, 2020).

2.3 Classifier

In this study, the drug–target prediction is modeled as the binary task. All the independent and interactive features learned in previous sections (Sections 2.2.1 and 2.2.2) are concatenated and fused to the fully connected dense layers in the downstream.

$$\hat{Y} = \delta(W_{out}[G_{drug}; L_{drug}; L_{prot}; G_{prot}] + b_{out}) \tag{11}$$

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

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

$$loss(\Theta) = \frac{1}{N}\sum_i -[y_i \cdot \log(\hat{y_i}) + (1 - y_i) \cdot \log(1 - \hat{y_i})] + \frac{\lambda}{2}\|\Theta\|_2^2 \tag{12}$$

where y_i is the real label, $\hat{y_i}$ is the predicted label, Θ is the set of all parameters in the end-to-end model (e.g. the combined networks of CNN, GAT, the bidirectional encoder–decoder block and classifier), N is the number of training samples, and λ is the L2 regularization coefficient.

3 Experiments

3.1 Experimental setup

In this study, the AUC, AUPR, precision and recall are chosen as the main metrics to evaluate the performance of the model. In the experiments, IIFDTI is implemented in Pytorch. We perform hyperparameter optimization on the DrugBank dataset. The dataset is divided into training set, validation set and test set according to the ratio of 8:1:1. We train the model on the training set and select the optimal hyperparameters based on the performance on the validation set, and finally perform model evaluation on the test set. We utilize grid search to determine the optimal hyperparameters of the model. The training loss is optimized by using AdamW (Loshchilov and Hutter, 2019) optimizer with a warmup strategy (He *et al.*, 2016). The selected hyperparameter settings of the model are listed in Supplementary Section S6.

3.2 Baseline methods

- Models based on independent features

DeepDTA (Öztürk *et al.*, 2018): This model consists of two 3-layer CNNs. It is originally designed for predicting binding affinity. Here, we change the last layer of its fully connected dense layers and the corresponding loss function, which makes the model suitable for DTIs predictions. We optimize the hyperparameters described from the work based on DrugBank dataset to ensure fairness.

DeepConv-DTI (Lee *et al.*, 2019): This model captures local residue patterns of targets by convolutions on various lengths of amino acid subsequences. It extracts features of drugs by using fully connected dense layers on drug fingerprints. Finally, the feature vectors of drugs and targets are concatenated to the fully connected dense layers for predicting DTIs. We follow the hyperparameter setting

described in the work because it has been optimized for data from DrugBank database.

- Models based on interactive features

MolTrans (Huang *et al.*, 2021): This model proposes a transformer-based bio-inspired molecular representation method. First, it constructs the sub-structural pattern mining algorithm. Then, the interaction map is built from drug sequences and target sequences. Finally, the CNN layer is applied on the interaction map to extract higher-order interactions for predicting DTIs. We follow the hyperparameter setting described in the work because it has been optimized for data from DrugBank database.

TransformerCPI (Chen *et al.*, 2020): This model includes a transformer-based neural network. Specifically, the features of target sequences are encoded to the decoder. At the same time, the decoder predicts DTIs based on the output of the encoder and the topological features of drugs extracted by graph neural network. We follow the hyperparameter setting described in the work.

The selected hyperparameter settings of baselines are listed in Supplementary Section S7.

3.3 Comparison of results

To comprehensively evaluate the prediction performance, according to the composition of different datasets, we run all the compared methods on BindingDB dataset and GPCR dataset ten times and get the averaged experimental results. On the other three datasets (DrugBank, Human and *C.elegans*), we get the averaged experimental results after 10 times repeated 5-fold cross-validation. The best results are highlighted in bold for each metric. The detailed dataset partition and training process are listed in Supplementary Section S1.

As shown in Table 2, for DrugBank dataset, IIFDTI has achieved more significant improvement compared to baselines. Compared with the second best method, IIFDTI achieves an improvement of 5.97% and 6.89% in terms of AUC and AUPR.

For small public datasets, such as Human and *C.elegans*, IIFDTI generally achieves better performance than baselines. IIFDTI is better than the second best method by 1.02% and 0.71% on Human and *C.elegans* datasets, respectively in terms of AUC. Similarly, IIFDTI is better by 0.92% and 0.71% in terms of AUPR.

For large-scale public datasets, on the BindingDB dataset, IIFDTI achieves an improvement of 1.07% and 1.17% in terms of AUC and AUPR. On the GPCR dataset, IIFDTI almost achieves the same performance as TransformerCPI and outperforms the other three baselines in terms of AUC and AUPR.

The above results indicate that the proposed IIFDTI achieves competitive results on various datasets compared to methods based only on independent features (DeepDTA, DeepConv-DTI) or methods based only on interactive features (MolTrans, TransformerCPI). When extracting independent features of drugs and targets, IIFDTI models the sequence information and graph structure information of drugs and targets based on CNN and GAT, while when extracting interactive features, the attention-based encoder–decoder can well preserve information between drug–target substructures. Combining the advantages of independent features and interactive features, IIFDTI achieves better results than baselines.

3.4 Ablation experiment

We conduct ablation experiments from two aspects, including component ablation and model design. The model structures of all ablation methods are listed in Supplementary Section S8.

In terms of the component ablation, we remove the relevant network structures to confirm their importance to the improvement of performances. IIFDTI (independent) indicates that the IIFDTI uses only independent features (Section 2.2.2). IIFDTI (interactive_a) indicates that the IIFDTI uses only interactive features (Section 2.2.1). IIFDTI (u_a) indicates that a unidirectional encoder–decoder structure is used for the interactive features in IIFDTI, in which drugs are inputted to the encoder and targets are inputted to the decoder. IIFDTI (u_b) indicates that a unidirectional encoder–decoder structure is used for the interactive features in IIFDTI, in which targets are inputted to the encoder and drugs are inputted to the decoder. From the results shown in Table 3, IIFDTI performs better than IIFDTI (u_a) and IIFDTI (u_b), which confirms that the use of bidirectional encoder–decoder structure is of great significance for

Table 2. The results on all the dataset: AUC, AUPR, precision, recall of the baselines and IIFDTI

Datasets	Methods	AUC (SD)	AUPR (SD)	Precision (SD)	Recall (SD)
DrugBank	DeepDTA	0.871 (0.003)	0.870 (0.004)	0.786 (0.015)	0.798 (0.020)
	DeepConv-DTI	0.836 (0.003)	0.831 (0.007)	0.736 (0.023)	0.767 (0.022)
	MolTrans	0.862 (0.007)	0.862 (0.010)	0.809 (0.006)	0.767 (0.015)
	TransformerCPI	0.865 (0.003)	0.868 (0.007)	0.774 (0.011)	0.821 (0.009)
	IIFDTI	**0.923 (0.005)**	**0.930 (0.006)**	**0.854 (0.021)**	**0.860 (0.022)**
Human	DeepDTA	0.972 (0.001)	0.973 (0.002)	0.938 (0.012)	0.935 (0.017)
	DeepConv-DTI	0.967 (0.002)	0.964 (0.004)	0.939 (0.018)	0.907 (0.023)
	MolTrans	0.974 (0.002)	0.976 (0.003)	**0.955 (0.012)**	0.933 (0.022)
	TransformerCPI	0.970 (0.006)	0.974 (0.005)	0.911 (0.021)	0.937 (0.011)
	IIFDTI	**0.984 (0.003)**	**0.985 (0.003)**	0.946 (0.017)	**0.947 (0.017)**
C.elegans	DeepDTA	0.983 (0.001)	0.984 (0.007)	0.970 (0.011)	0.960 (0.010)
	DeepConv-DTI	0.983 (0.002)	0.985 (0.001)	0.954 (0.006)	0.936 (0.008)
	MolTrans	0.982 (0.003)	0.982 (0.003)	**0.971 (0.007)**	0.963 (0.012)
	TransformerCPI	0.984 (0.002)	0.983 (0.003)	0.949 (0.011)	0.948 (0.012)
	IIFDTI	**0.991 (0.002)**	**0.992 (0.003)**	0.954 (0.010)	**0.971 (0.011)**
BindingDB	DeepDTA	0.934 (0.007)	0.934 (0.008)	0.858 (0.021)	0.860 (0.023)
	DeepConv-DTI	0.922 (0.003)	0.921 (0.004)	0.835 (0.024)	0.846 (0.031)
	MolTrans	0.899 (0.006)	0.897 (0.010)	0.826 (0.021)	0.768 (0.019)
	TransformerCPI	0.933 (0.011)	0.934 (0.015)	0.840 (0.023)	**0.891 (0.022)**
	IIFDTI	**0.944 (0.003)**	**0.945 (0.004)**	**0.879 (0.011)**	0.873 (0.013)
GPCR	DeepDTA	0.776 (0.006)	0.762 (0.015)	0.713 (0.014)	0.712 (0.015)
	DeepConv-DTI	0.752 (0.011)	0.685 (0.010)	0.695 (0.020)	0.713 (0.021)
	MolTrans	0.807 (0.004)	0.788 (0.009)	0.699 (0.007)	0.762 (0.014)
	TransformerCPI	0.842 (0.007)	0.837 (0.010)	0.755 (0.013)	**0.796 (0.015)**
	IIFDTI	**0.845 (0.008)**	**0.842 (0.007)**	**0.766 (0.009)**	0.783 (0.017)

The best results are highlighted in bold for each metric.

Table 3. The results on DrugBank: AUC, AUPR, precision and recall of IIFDTI (independent), IIFDTI (interactive_a), IIFDTI (u_a), IIFDTI (u_b) and IIFDTI

Methods	Number of parameters	AUC (SD)	AUPR (SD)	Precision (SD)	Recall (SD)
IIFDTI (independent)	15.9 M	0.859 (0.002)	0.855 (0.006)	0.794 (0.016)	0.787 (0.019)
IIFDTI (interactive_a)	17.5 M	0.867 (0.005)	0.886 (0.005)	0.805 (0.018)	0.781 (0.019)
IIFDTI (u_a)	24.2 M	0.919 (0.004)	0.924 (0.002)	**0.867 (0.011)**	0.833 (0.023)
IIFDTI (u_b)	24.2 M	0.913 (0.004)	0.918 (0.007)	0.858 (0.018)	0.823 (0.023)
IIFDTI	29.8 M	**0.923 (0.005)**	**0.930 (0.006)**	0.854 (0.021)	**0.860 (0.022)**

The best results are highlighted in bold for each metric.

Table 4. The results on DrugBank: AUC, AUPR, precision and recall of IIFDTI (independent), IIFDTI (interactive_b), IIFDTI (interactive_c), IIFDTI (independent+interactive_b), IIFDTI (independent+interactive_c) and IIFDTI

Methods	Number of parameters	AUC (SD)	AUPR(SD)	Precision (SD)	Recall (SD)
IIFDTI (independent)	15.9 M	0.859 (0.002)	0.855 (0.006)	0.794 (0.016)	0.787 (0.019)
IIFDTI (interactive_b)	24.8 M	0.875 (0.008)	0.889 (0.011)	0.831 (0.023)	0.783 (0.015)
IIFDTI (interactive_c)	228.0 M	0.852 (0.009)	0.846 (0.012)	0.787 (0.017)	0.772 (0.012)
IIFDTI (independent+ interactive_b)	29.2 M	0.905 (0.010)	0.920 (0.010)	**0.929 (0.022)**	0.681 (0.008)
IIFDTI (independent+ interactive_c)	228.1 M	0.869 (0.009)	0.874 (0.010)	0.779 (0.013)	0.807 (0.015)
IIFDTI	29.8 M	**0.923 (0.005)**	**0.930 (0.006)**	0.854 (0.021)	**0.860 (0.022)**

The best results are highlighted in bold for each metric.

Table 5. The results on Davis: AUC, AUPR, precision and recall of baselines and IIFDTI

Methods	AUC (SD)	AUPR (SD)	Precision (SD)	Recall (SD)
DeepDTA	0.906 (0.002)	0.862 (0.004)	0.788 (0.022)	0.743 (0.019)
DeepConv-DTI	0.901 (0.002)	0.821 (0.005)	0.762 (0.021)	0.706 (0.025)
MolTrans	0.892 (0.004)	0.794 (0.008)	0.784 (0.014)	0.623 (0.012)
TransformerCPI	0.846 (0.001)	0.748 (0.001)	0.689 (0.002)	0.607 (0.004)
IIFDTI	**0.911 (0.002)**	**0.888 (0.003)**	**0.805 (0.001)**	**0.823 (0.003)**

The best results are highlighted in bold for each metric.

the performance improvement of IIFDTI. IIFDTI also performs better than IIFDTI (independent) and IIFDTI (interactive_a) on DrugBank dataset, which indicates that the combination of interactive features from the bidirectional encoder–decoder structure and independent features from GAT and CNN can improve the performance of the proposed model. Similar results on other benchmark datasets are available in Supplementary Section S9.

In terms of the model design, the modules of our model are replaced by other existing structures to confirm the conclusion that combining independent and interactive features is effective. IIFDTI (interactive_b) contains only interactive features, which replace the original encoders for drug–target pairs with GAT and CNN structures. IIFDTI (interactive_c) contains only interactive features, which replaces the bidirectional encoder–decoder structure with a different interactive feature extraction module proposed by the previous work (Huang *et al.*, 2021). IIFDTI (independent+interactive_b) is composed of IIFDTI (independent) and IIFDTI (interactive_b), and IIFDTI (independent+interactive_c) is composed of IIFDTI (independent) and IIFDTI (interactive_c). The results show that IIFDTI (independent+interactive_b) outperforms IIFDTI (independent) and IIFDTI (interactive_b), IIFDTI (independent+interactive_c) outperforms IIFDTI (independent) and IIFDTI (interactive_c). From the perspective of model design, the ablation experiments indicate that combining independent features and interactive features is effective. Furthermore, IIFDTI also outperforms IIFDTI (independent+interactive_b) and IIFDTI (independent+interactive_c), and it indicates that the combined GAT, CNN and transformer structure enhances the performances of IIFDTI. This may be due to the complementarity of features extracted by different feature extractors. Similar results on other benchmark datasets are available in Supplementary Section S10.

Table 6. Summary of the training set and test set

Datasets	Targets	Drugs	Interactions	Positive	Negative
Training set	1606	1471	12 624	6312	6312
Test set	1326	1144	4624	949	3675

Overall, the ablation studies indicate that fusing independent features and interactive features and choosing appropriate model components are effective for improving the performance of our framework.

3.5 Imbalanced dataset and independent test set experiments

In order to test the performances of IIFDTI on imbalanced dataset, we also construct the imbalanced benchmark dataset, Davis (Davis *et al.*, 2011). Davis records wet lab assay values measuring binding affinities among drugs and targets. We divide all samples into positive and negative samples by the threshold 5.0 following previous works (Davis *et al.*, 2011; Öztürk *et al.*, 2018). The details of the Davis dataset are listed in Supplementary Section S11. Finally, the Davis dataset contains 7320 positive samples and 18 452 negative samples between 379 targets and 68 drugs, and the ratio of positive and negative samples is about 1:2.5. As shown in Table 5, IIFDTI outperforms other methods, and IIFDTI is better than the second best method by 0.50% and 3.01% in terms of AUC and AUPR, respectively. The experimental results indicate that IIFDTI is still reliable under imbalanced dataset training.

Table 7. The results of independent test set experiment: AUC, AUPR, precision and recall of baselines and IIFDTI

Methods	AUC (SD)	AUPR (SD)	Precision (SD)	Recall (SD)
DeepDTA	0.784 (0.004)	0.519 (0.007)	0.450 (0.006)	0.635 (0.010)
DeepConv-DTI	0.782 (0.005)	0.472 (0.005)	**0.459 (0.013)**	0.626 (0.016)
MolTrans	0.501 (0.010)	0.203 (0.006)	0.183 (0.032)	0.417 (0.015)
TransformerCPI	0.782 (0.005)	0.500 (0.015)	0.427 (0.008)	0.660 (0.007)
IIFDTI	**0.797 (0.004)**	**0.527 (0.009)**	0.436 (0.002)	**0.679 (0.008)**

The best results are highlighted in bold for each metric.

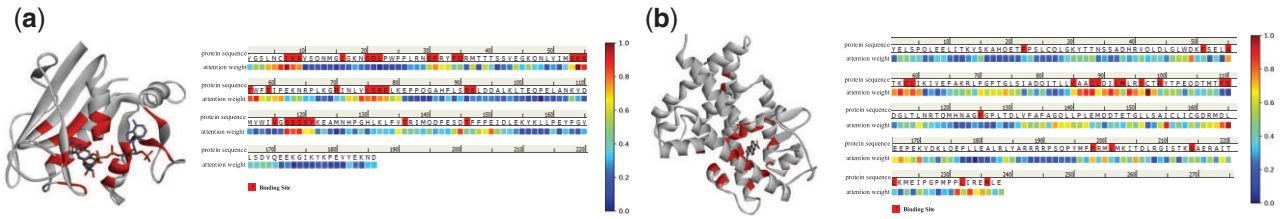

Fig. 3. Visualizations of attention weights. (**a**) attention weight of *Dihydrofolate reductase* (PDB:3FS6). (**b**) Attention weight of *Retinoic acid receptor Y* (PDB:6FX0)

Table 8. The predicted candidate targets for new drug *Diacerein*

Rank	Target name	Target Uniprot ID	Source
1	*Prostaglandin G/H synthase 1*	P23219	Unknown
2	*Cyclooxygenase-2*	P35354	Unknown
3	***Cytochrome P450 3A4***	**P08684**	**PubMed: 18814214**
4	*GMP-PDE gamma*	P18545	Unknown
5	*Calstabin-1*	P62942	Unknown
6	***Cholesterol 25-hydroxylase***	**P11712**	**PubMed: 18814214**
7	*Lactoylglutathione lyase*	Q04760	Unknown
8	***Cytochrome P450 1A2***	**P05177**	**PubMed: 18814214**
9	*Dihydrofolate reductase*	P00374	Unknown
10	***CYPIID6***	**P10635**	**PubMed:18814214**

The best results are highlighted in bold for each metric.

To better evaluate the generalization ability of the model, we simulate the real experimental environment by collecting the training set and the test set on the time series. As shown in Table 6, the approved data of DrugBank release version 5.0.3 (before October 2016) is used as training set. Then, we filter the newly approved data after October 2016 as an independent test set with ensuring that the drugs or targets in the test set do not appear in the training set.

As shown in Table 7, we get the experimental results with the cross-validation for 10 times. Compared with other methods, the improvement of IIFDTI is significant. IIFDTI achieves an improvement of 1.65% and 1.54% in terms of AUC and AUPR compared with the second best method. The results indicate that IIFDTI can deal with problems for an independent test set in a real data environment.

3.6 Attention interpretation

In order to better analyze what the models have learned and what they are based on to make predictions, we analyze the attention weights learned by models, whose visualizations and interpretations can greatly help us to design future models. Based on the attention mechanism in encoder–decoder framework, we can simply understand the underlying mechanism based on the attention weights of drug sequences and target sequences.

As shown in Figure 3, the *Dihydrofolate reductase* (PDB:3FS6) and *Retinoic acid receptor Y* (PDB:6FX0) with their corresponding actives are selected as examples. For each example, the left side shows the 3D protein structures, in which the binding sites are highlighted in red. The right side shows the amino acid sequence of the protein and the model-learned attention weights for each residue, visualized with a heatmap to better distinguish them. These residues that are marked redder indicate higher weights, and the actual binding sites on the amino acid sequence are marked in dark red. As shown in Figure 3a, the sequence fragments with higher attention weight visualized by heatmap cover multiple binding site positions, e.g. the 5th to 9th, 53rd to 57th, 74th to 78th and 117th to 121st positions in the amino acid sequence. Similarly, as shown in Figure 3b, sequence fragments with higher attention weight (e.g. 55th to 60th, 86th to 98th and 231st to 233rd) also cover many known binding sites. From the results, IIFDTI can almost successfully localize the binding site of the ligand to these high-weight regions. It indicates that our model has learned the knowledge insights and biological significances from large amounts of data.

3.7 Case study

In order to further verify the effectiveness of the proposed model, we apply IIFDTI for *de novo* predictions on two important drugs and targets [*Diacerein* (DrugBank ID: DB11994) and *Aspartate aminotransferase* (Uniprot ID: Q2TU84)], respectively. In this section, we use the pre-trained model to predict the interactive probabilities between them and known drugs or targets in the datasets. For the prediction results, we sort the candidate drugs (or targets) according to their prediction scores. Finally, the predicted interactions are verified through the DrugBank database (Wishart *et al.*, 2006).

As shown in Table 8, the top 10 candidate targets for *Diacerein* predicted by IIFDTI are listed and there are 11 validated targets of

Table 9. The predicted candidate drugs for new target *Aspartate aminotransferase*

Rank	Drug name	DrugBank ID	Source
1	*Citric acid*	DB04272	Unknown
2	*Adenosine triphosphate*	DB00171	Unknown
3	*Phosphoaminophosphonic Acid-Adenylate Ester*	DB04395	Unknown
4	***Pyridoxal phosphate***	**DB00114**	**PubMed: 240513**
5	***Glutamic acid***	**DB00142**	**PubMed: 24121043**
6	*Adenosine 3',5'-diphosphate*	DB01812	Unknown
7	*Adenylylsulfate*	DB03708	Unknown
8	*Pidolic acid*	DB03088	Unknown
9	*Acetohydroxamic acid*	DB00551	Unknown
10	***Aspartic acid***	**DB00128**	**PubMed: 26232224**

The best results are highlighted in bold for each metric.

all 2176 targets. From the result, we can find that 4 out of 10 targets are predicted successfully (marked in bold). Similarly, Table 9 shows the predicted candidate drugs for the new target *Aspartate aminotransferase*, for which there are 3 validated drugs of all 4400 drugs. All 3 candidate drugs are predicted successfully in the top 10 (marked in bold).

Overall, the above two cases (i.e. *Diacerein* and *Aspartate aminotransferase*) are considered challenging where they need to predict candidate drugs (or targets) for new targets (or drugs) precisely. The results demonstrate that the IIFDTI has the ability to accurately predict interactive candidates for unknown drugs and targets from large sample sets, which has a great significance for drug screening and drug reposition.

4 Conclusion

Identifying the potential drug–target interactions is a crucial task in drug discovery and drug reposition. Although existing research has achieved great success, improving the performance of DTI predictions still remains a major challenge. In this study, an end-to-end deep learning model named IIFDTI based on the fusion of independent and interactive features is proposed for predicting DTIs. Compared with existing state-of-the-art models on several benchmark datasets, the computational results show that IIFDTI can achieve greater performance than baselines. Simultaneously, the mapped visualizations of attention weights learned by IIFDTI provide meaningful guidance for interpretations of biological significance and the prediction optimization of DTIs.

Predicting DTIs is an extremely complex and difficult problem. It is clear that IIFDTI still has room for further improvement. In this study, only one type of data, such as amino acid sequences for targets and SMILES strings for drugs, is learned as input. In practice, drugs and targets have different representations from different levels. The modeling and fusion of the information provide a possibility to improve the performance of predicting DTIs. In future work, we will focus on the fusion of multi-view representations in deep learning-based models.

Funding

The authors would like to express their gratitude for the support from the National Key Research and Development Program of China [no. 2021YFF1201200], the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization [no. U1909208], the National Natural Science Foundation of China (No. 61972423, and No. 62072473), 111 Project [no. B18059] and Hunan Provincial Science and Technology Program [no. 2018WK4001].

Conflict of Interest: none declared.

References

Asgari,E. and Mofrad,M.R. (2015) Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS One*, **10**, e0141287.

Beltagy,I. *et al.* (2020). Longformer: the long-document transformer. arXiv preprint arXiv:2004.05150. https://doi.org/10.48550/arXiv.2004.05150.

Bleakley,K. and Yamanishi,Y. (2009) Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics*, **25**, 2397–2403.

Chan,W.K. *et al.* (2015) Glass: a comprehensive database for experimentally validated GPCR-ligand associations. *Bioinformatics*, **31**, 3035–3042.

Chen,L. *et al.* (2020) TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, **36**, 4406–4414.

Cheng,F. *et al.* (2012) Prediction of chemical–protein interactions: multitarget-QSAR versus computational chemogenomic methods. *Mol. Biosyst.*, **8**, 2373–2384.

Dai,Z. *et al.* (2019). Transformer-xl: Attentive language models beyond a fixed-length context. In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, Florence, Italy, July 28–August 2, 2019*.

Dauphin,Y.N. *et al.* (2017). Language modeling with gated convolutional networks. In: *Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017*.

Davis,M.I. *et al.* (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1046–1051.

Devlin,J. *et al.* (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, June 2–7, 2019*.

Ezzat,A. *et al.* (2019) Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief. Bioinform.*, **20**, 1337–1357.

Gao,K.Y. *et al.* (2018). Interpretable drug target prediction using deep neural representation. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, July 13–19, 2018*.

Gilson,M.K. *et al.* (2016) Bindingdb in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–D1053.

He,K. *et al.* (2016). Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, June 27–30, 2016*.

Huang,K. *et al.* (2020) DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics*, **36**, 5545–5547.

Huang,K. *et al.* (2021) MolTrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, **37**, 830–836.

Ioffe,S. and Szegedy,C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015*.

Karimi,M. *et al.* (2019) DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, **35**, 3329–3338.

Lee,I. *et al.* (2019) DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.*, **15**, e1007129.

Liu,H. *et al.* (2015) Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics*, **31**, i221–i229.

Loshchilov,I. and Hutter,F. (2019). Decoupled weight decay regularization. In: *7th International Conference on Learning Representations, New Orleans, LA, USA, May 6–9, 2019*.

Mikolov,T. *et al.* (2013). Efficient estimation of word representations in vector space. In: *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2–4, 2013*.

Mohamed,S.K. *et al.* (2020) Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, **36**, 603–610.

Nguyen,T. *et al.* (2021) GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, **37**, 1140–1147.

Olayan,R.S. *et al.* (2018) DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics*, **34**, 1164–1173.

Öztürk,H. *et al.* (2018) DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, **34**, i821–i829.

Quan,Z. *et al.* (2018). A system for learning atoms based on long short-term memory recurrent neural networks. In: *IEEE International Conference on Bioinformatics and Biomedicine, Madrid, Spain, December 3–6, 2018*.

Rifaioglu,A.S. *et al.* (2021) MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. *Bioinformatics*, **37**, 693–704.

Schenone,M. *et al.* (2013) Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.*, **9**, 232–240.

Tsubaki,M. *et al.* (2019) Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, **35**, 309–318.

Vaswani,A. *et al.* (2017). Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems* (NeurIPS 2017). Long Beach, CA, USA, pp. 5998–6008.

Velickovic,P. *et al.* (2018). Graph attention networks. In: *6th International Conference on Learning Representations, Vancouver, BC, Canada, April 30–May 3, 2018*.

Wan,F. *et al.* (2019) NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*, **35**, 104–111.

Wang,F. *et al.* (2011) Computational screening for active compounds targeting protein sequences: methodology and experimental validation. *J. Chem. Inf. Model.*, **51**, 2821–2828.

Wang,Y. and Zeng,J. (2013) Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics*, **29**, i126–i134.

Wishart,D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.

Yuan,Q. *et al.* (2016) DrugE-Rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*, **32**, i18–i27.

Zheng,S. *et al.* (2020) Predicting drug–protein interaction using quasi-visual question answering system. *Nat. Mach. Intell.*, **2**, 134–140.

Zong,N. *et al.* (2017) Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations. *Bioinformatics*, **33**, 2337–2344.