BMC Biology

RESEARCH

Open Access

# SMFF-DTA: using a sequential multi-feature fusion method with multiple attention mechanisms to predict drug-target binding affinity

Xun Wang[1,2], Zhijun Xia[1,2], Runqiu Feng[1,2], Tongyu Han[1,2], Hanyu Wang[1,2], Wenqian Yu[1,2] and Xingguang Wang[3*]

## Abstract

**Background** Drug-target binding affinity (DTA) prediction can accelerate the drug screening process, and deep learning techniques have been used in all facets of drug research. Affinity prediction based on deep learning methods has proven crucial to drug discovery, design, and reuse. Among these, the sequence-based approach using 1D sequences of drugs and targets as inputs typically results in the loss of structural information, whereas the structure-based method frequently results in increased computing costs due to the intricate structure of the molecule graph.

**Results** We propose a sequential multifeature fusion method (SMFF-DTA) to achieve efficient and accurate prediction. SMFF-DTA uses sequential methods to represent the structural information and physicochemical properties of drugs and targets and introduces multiple attention blocks to capture interaction features closely.

**Conclusions** As demonstrated by our extensive studies, SMFF-DTA outperforms the other methods in terms of various metrics, showing its advantages and effectiveness as a drug-target binding affinity predictor.

**Keywords** Drug-target binding affinity, Multifeature, Deep learning, Attention

## Background

Drug discovery is a time-consuming and expensive process that typically costs 10–15 years and over $200 million to launch a medicine effectively [1]. In practice, drugs usually function as ligands, interacting with target proteins to exert their specific effects. The efficacy of a drug typically depends on how strongly it binds to its target protein, and determining this process is critical to uncovering the mechanism of drug action. Binding affinity represents the interaction strength between drug–target pairs, and its prediction via experimental or computational methods is called drug–target binding affinity (DTA) prediction. Drugs with strong interactions are often selected as potential active compounds according to their predicted affinity values [2], providing candidate drugs for subsequent biological wet experiment verification [3] and promoting drug discovery, design and reuse. Binding affinity clearly serves as an essential indicator of the strength of drug–target interactions, which constitute the core of drug research and development [4].

*Correspondence:
Xingguang Wang
xingguang-wang@163.com
[1] Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Changjiang West Road, Qingdao 266580, Shandong, China
[2] Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, Qingdao, Shandong 266580, China
[3] Shandong Provincial Hospital, Jingwu Weiqi Road, Jinan 250021, Shandong, China

Early approaches for DTA prediction are traditionally achieved via wet experiments, which tend to be reliable and accurate, but the process is demanding and costly. With the advent of computer-aided drug design techniques, prediction methods based on physical or molecular docking have been developed. However, physics-based prediction methods, such as free energy simulation and molecular dynamics simulation, often result in considerable computational overhead in practical applications. The molecular docking method uses a search algorithm to predict 3D complexes formed by drug–target pairs [5], but the results are always inaccurate. As a data-driven approach, machine learning models are excellent at processing low-dimensional data with minimal computational costs and high interpretability [6]. For example, KronRLS [7] predicts binding affinity via the Kronecker regularized least square method, whereas SimBoost [8] constructs features of drugs, targets and drug–target pairs via gradient boosting machines to make predictions and generate prediction intervals. However, machine learning-based methods rely heavily on hand-extracted features, which requires researchers to have extensive domain knowledge [9].

With advances in deep learning technology, deep learning-based approaches are increasingly being applied to all stages of drug development [10]. Since it is able to automatically generate feature representations from raw inputs without the need for any relevant domain knowledge [11], this makes it much easier to process biological data with extremely complex internal structures. Deep learning-based methods can accelerate the process of drug screening, thus greatly reducing development costs and time. Drugs and targets are intrinsically complex biological data that contain abundant feature information. The most widely used features are 1D sequences, including simplified molecular linear input specification (SMILES) of drugs and amino acid sequences of targets, from which deep learning methods such as convolutional neural networks (CNNs) [11], recurrent neural networks (RNNs) [12], transformers [13] and others are then employed to extract implicit features. For example, DeepDTA [14] uses a CNN to extract drug and target features separately through successive 1D convolutional layers. GANsDTA [15] employs generative adversarial networks (GANs) to extract features in an unsupervised manner. TF-DTA [16] uses the encoder modules of the transformer and multilayer CNNs to obtain better protein and drug representations, respectively. However, such sequence representations often ignore important structural information. Therefore, models that consider the molecular graph of drugs and distance maps or contact maps of targets as inputs, which extract structural features through various graph neural networks (such as GCN [17] and GAT [18]), can be developed. For example, when a molecule graph is used as the drug input, LLMDTA [19] focuses more on the intricate feature of the drug component and attempts to integrate the drug structure into the sequence-only method. MultiDTA [20] introduces both drug molecular graphs and target contact maps to provide more comprehensive features through the fusion of multimodal information. However, when dealing with large-scale graphs, it is difficult for a graph neural network to compute as efficiently as sequential models because of the complex relationships between nodes and edges. Furthermore, how to effectively fuse multimodal information remains a challenge for methods that use both sequence and graph inputs.

Considering the aforementioned issues, we propose a novel deep learning model, SMFF-DTA, to predict drug-target binding affinity. SMFF-DTA represents the structural information of drugs and targets in sequential ways and innovates encoding methods for the physicochemical properties of drug atoms and amino acid residues. Additionally, we construct a feature encoder to implement feature extraction in both local and global modes and design a multiple attention block to extract crucial interaction features between drugs and targets in both direct and indirect ways. On the basis of the experimental results, SMFF-DTA has demonstrated its ability to accurately predict drug-target binding affinity, as it outperforms other novel advanced methods.

## Results and discussion

### Performance evaluation

We contrast the seven state-of-the-art methods discussed in the Introduction section with our proposed model SMFF-DTA. As seen from the results in Table 1, SMFF-DTA has achieved better performance in both Davis and KIBA. Compared with the second-best results, the evaluation indices MSE, $R_m^2$, and CI are improved by 2%, 1.6%, and 0.4%, respectively, in Davis and by 0.5%, 1.2%, and 0.4%, respectively, in KIBA. SMFF-DTA is a potent approach for affiinity prediction, as evidenced by its outstanding performance in terms of error, correlation, and accuracy.

### Ablation experiments

To illustrate the necessity of multifeature inputs, we test the effects of different kinds of inputs on model performance. More precisely, sequences include drug SMILES and target amino acid sequences, Sequence+Structure adds drug fingerprints and target secondary structures, and Sequence+Structure+Properties adds the physicochemical properties of drug atoms and target residues. As shown in Fig. 1a, multifeature inputs can increase model

**Table 1** Performance comparison of SMFF-DTA and state-of-the-art methods on benchmark datasets

| Method | Davis | | | KIBA | | |
|---|---|---|---|---|---|---|
| | MSE | $R_m^2$ | CI | MSE | $R_m^2$ | CI |
| KronRLS (2015) | 0.379 | 0.407 | 0.871 | 0.411 | 0.342 | 0.782 |
| SimBoost (2017) | 0.282 | 0.644 | 0.872 | 0.222 | 0.629 | 0.836 |
| DeepDTA (2018) | 0.261 | 0.630 | 0.878 | 0.194 | 0.673 | 0.863 |
| GANsDTA (2020) | 0.276 | 0.653 | 0.881 | 0.224 | 0.675 | 0.866 |
| TF-DTA (2023) | 0.231 | 0.670 | 0.886 | 0.177 | 0.734 | 0.877 |
| MultiDTA (2024) | 0.231 | 0.694 | *0.893* | *0.156* | 0.761 | *0.890* |
| LLMDTA (2024) | *0.226* | *0.717* | 0.884 | 0.162 | *0.768* | 0.872 |
| SMFF-DTA (our) | **0.206** | **0.733** | **0.897** | **0.151** | **0.780** | **0.894** |

The best results in the metrics are highlighted in bold, and the second-best results are italicized

performance; thus, achieving comprehensive input representations is necessary.

To demonstrate the effectiveness of capturing interaction information in both direct and indirect ways, we conduct ablation experiments on cross-attention, interaction blocks, and multiple attention blocks. Figure 1b clearly shows that model performance will suffer if any of the interaction capture parts are eliminated. This suggests that direct and indirect approaches to capture interactions can work in conjunction and are both capable of efficiently extracting interaction features.

To prove the effectiveness of our proposed encoding method for the physicochemical properties of drug atoms, we compare three encoding methods in total. Method 1 does not consider the order of atoms in SMILES and directly fills the first x row of the atom feature matrix, where x is the number of atoms in SMILES. Methods 2 and 3 take the order in account, only with the difference in the filling of nonatomic bits (such as bonds and ring structures), where Method 2 considers only atom bits and puts the nonatomic bits with zero, whereas Method 3 fills the nonatomic bits with features of the previous atom. The results in Fig. 1c show that Method 3 achieves the best performance, which indicates that the filling of nonatomic bits can consider the features of adjacent nodes to a certain degree and effectively expresses the physicochemical properties of drug atoms in a sequential way.

### Case study

Nine compounds are chosen at random from the PDBbind [21] database, and we ensure that the protein–ligand pairs are not present in Davis or KIBA. Besides, due to the larger data volume and more interactions in the KIBA dataset, we decide to use the model trained on KIBA to predict the affinity of these samples. As demonstrated by the findings in Table 2, 8 out of 9 samples are consistent with the ranking of PDBbind experimental values. This, to some extent, proves the generalizability and applicability of our model, which is capable of accurately extracting the features of unseen drugs and targets.

We further analyze the mis-pridected sample (PDBid: 1g2k). From our point of view, the most likely reason for the prediction error is because that the number of short protein sequences (less than 200) in KIBA is 0, largely leading to the insufficient training for short sequences. However, the length of the protein sequence normally is not set, so this guide us to investigate further in DTA prediction, trying to figure out how to better account for the various protein sequence lengths to improve the universality of the model.

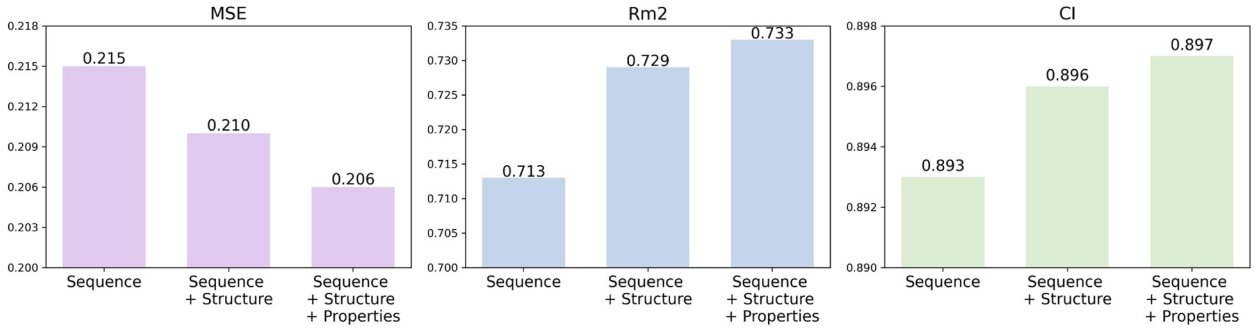
## Conclusions

We propose a drug-target binding affinity prediction model, SMFF-DTA, on the basis of multiple sequential features and multiple types of attention, which provides a sequential representation of the structural information and physicochemical properties of drugs and targets and encodes the physicochemical properties of drug atoms and amino acid residues in a novel manner. Moreover, directly and indirectly capturing the important interaction features simultaneously makes our model straightforward to interpret. The experimental outcomes demonstrate that SMFF-DTA has improved accuracy, correlation, and error and has a strong capacity for generalization, making it a useful technique for drug reuse and screening.

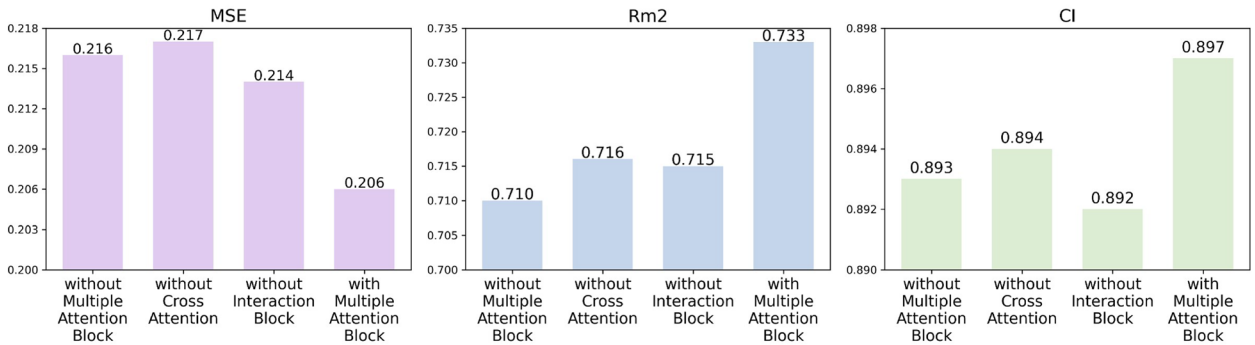
## Methods

### Datasets

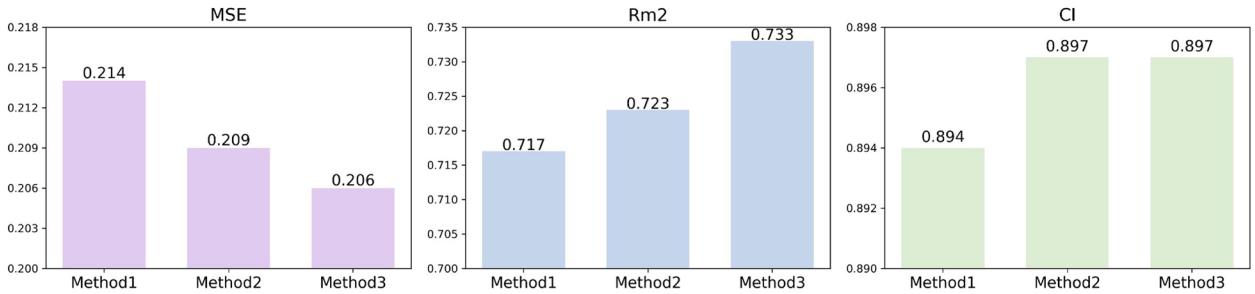
To evaluate our model, we conduct experiments on two widely used high-quality public datasets: Davis [22] and KIBA [23]. The basic state of the above datasets is presented in Table 3.




(a) Ablation experiments on different kinds of inputs on model performance. "Sequences" include drug SMILES and target amino acid sequences, "Sequence + Structure" adds drug fingerprints and target secondary structures, and "Sequence + Structure + Properties" adds the physicochemical properties of drug atoms and target residues.




(b) Ablation experiments on the effectiveness of capturing interaction information in both direct and indirect ways on model performance. We conduct ablation experiments on cross-attention, interaction blocks, and multiple attention blocks.




(c) Ablation experiments on different encoding methods for the physicochemical properties of drug atoms on model performance. Method1 does not consider the order of atoms in SMILES while Methods 2 and 3 take the order in account. But Method 2 considers only atom bits and puts the nonatomic bits with zero, whereas Method 3 fills the nonatomic bits with features of the previous atom. Method 3 is the innovative encoding approach we have proposed to serialize the physicochemical properties of drug atoms.

**Fig. 1** Details of ablation results. **a** Ablation experiments on different kinds of inputs. **b** Ablation experiments on Multiple Attention Block. **c** Ablation experiments on encoding methods for physicochemical properties of drug atoms

Davis contains 68 drugs and 442 targets, forming a total of 30,056 drug-target samples. Its affinity data are expressed by the kinase dissociation constant (Kd, units:M), which is usually converted into the log space form (pKd) to reduce variance, ranging from 5.0 to 10.8. KIBA contains 2111 drugs and 229 targets, resulting in 118,254 drug-target samples. Its affinity data, which is called KIBA score, ranges from 0 to 17.2. KIBA score unifies drug-target interaction data under various experimental conditions into a dimensionless comprehensive score by integrating data from multiple sources of biological activity (such as the inhibition constant (Ki, units: M), the half maximum inhibitory concentration (IC50, units: M) and Kd), using statistical weighting techniques and implementing standardized transformations via negative logarithm processing. The detailed information for drugs and targets in Davis and KIBA is displayed in Fig. 2. With the goal of covering as much data

**Table 2** Case study on samples from the PDBbind dataset

| PDBid | PDB value (pKd) | Predicted value (KIBA score) |
|---|---|---|
| 3b65 | 9.27 | 11.61 |
| 4eo8 | 8.15 | 11.40 |
| 1 g2k | 7.96 | 11.63 |
| 4 djv | 6.72 | 11.32 |
| 2iwx | 6.68 | 11.28 |
| 2fxs | 6.06 | 11.13 |
| 2xca | 5.60 | 11.00 |
| 3r88 | 4.82 | 10.90 |
| 1a30 | 4.30 | 10.86 |

**Table 3** Statistics of the benchmarking datasets

| Dataset | Drugs | Proteins | Interactions |
|---|---|---|---|
| Davis | 68 | 442 | 30,056 |
| KIBA | 2111 | 229 | 118,254 |

as possible and maintaining model performance, we ultimately set the target sequence length and drug SMILES length to 1200 (86% in Davis, 90% in KIBA) and 100 (100% in Davis, 97% in KIBA), respectively, on the basis of the distribution of the data. We set the number of drug atoms to 100 to facilitate the subsequent feature fusion process for the above reasons.

## Input representation

The input of SMFF-DTA is composed of 2 main parts: drugs and targets. To obtain an extensive feature input that covers reasonably thorough information, we explicitly extract the structural information of drugs and targets and the physicochemical properties of atoms and residues in addition to their 1D sequence representation.

### *Drug input*

The simplified molecular input line entry system (SMILES) [24] is a chemical specification that uses ASCII strings to describe the molecular structure, providing feature information such as atomic type, bond type and stereochemistry. Through label encoding, we create a dictionary in which each character of SMILES is assigned a distinct integer, totaling 64 integers. Hence, the sequence feature of drug input is shown by the drug SMILES matrix $F_{smi} \in \mathbb{R}^{bs \times 100}$, where bs is the batch size. Additionally, Morgan fingerprints encode topological features of drug molecules, capturing higher-level features such as substructure information (functional groups, ring structures, etc.), atom and bond information (atomic type, bond type, etc.), and topological structure (branching structure, molecular graph connectivity, etc.). In this way, the structural

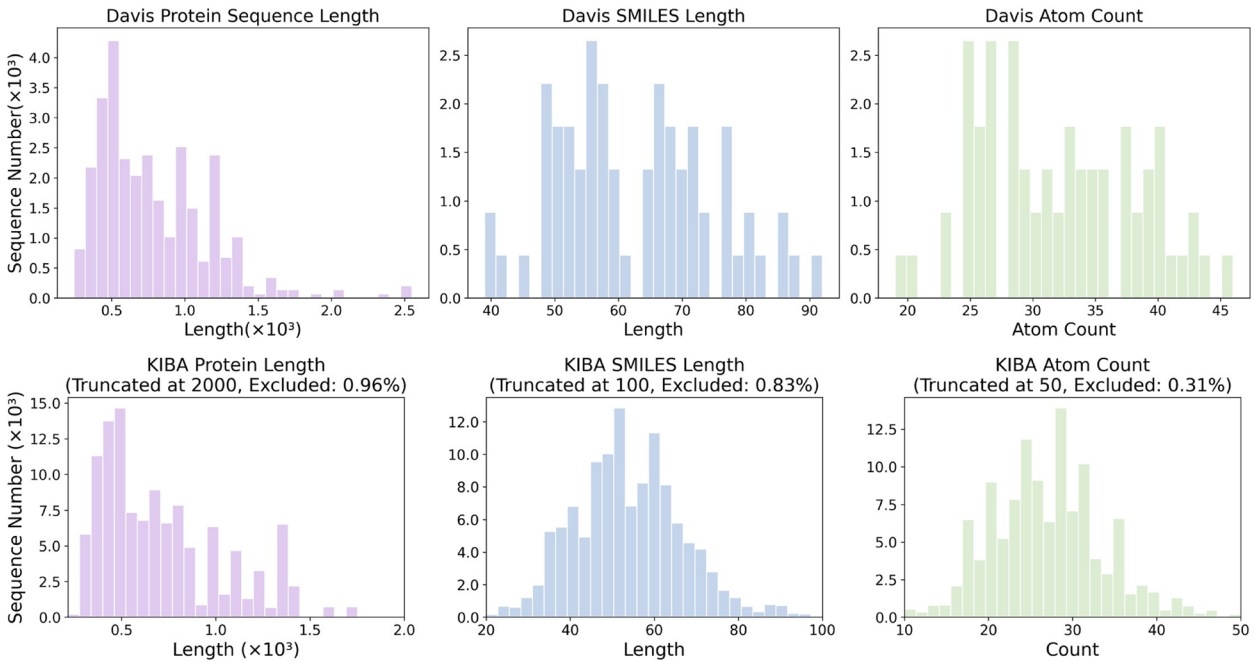



**Fig. 2** Lengths of the target sequence, lengths of drug SMILES, and counts of drug atoms in Davis and KIBA. For KIBA dataset, we only plot the distribution of data within the specified length (protein sequence length: 2000, SMILES string length: 100, drug atom number count: 100)

information of drugs can be conveyed implicitly, and the Morgan technique can remove some ambiguous atomic identifiers [25]. We use RDKit [26] to convert the SMILES strings to Morgan fingerprints [27], setting the iteration radius to 2, thus obtaining the binary fingerprint matrix $F_{fp} \in \mathbb{R}^{bs\times 1024}$. Combining Morgan fingerprints with drug SMILES can improve the model's generalizability and reduce the model's sensitivity to single inputs. However, it is worth noting that neither the SMILES nor Morgan fingerprints mentioned above can represent the specific property information of drug atoms. Given this, we utilize RDKit to extract the related physicochemical properties from SMILES and represent it as a 38D one-hot atom feature vector. The atomic feature information we use is displayed in Table 4.

Moreover, we propose a novel encoding strategy to depict the physicochemical properties of drug atoms, as illustrated in Fig. 3. The position of each atom in SMILES is noted, and the atomic bits in SMILES are filled with physicochemical features of corresponding atoms, whereas the nonatomic bits (such as bonds) are filled with physicochemical features of preceding atoms to form the atom feature matrix $F_{atom} \in \mathbb{R}^{bs\times 100\times 38}$. By this means, we can take the physicochemical properties of neighboring atoms into account through the following feature extraction procedure, simulating the adjacency relationship such as the drug graph does to some extent.

### *Target input*

The amino acid composition of the entire target protein chain contains extensive sophisticated protein features, including functional, evolutionary and other implicit features. The protein secondary structure refers to the main chain of the protein peptide in the spatial arrangement of the atoms, providing information on the spatial structure [28]. We obtain the secondary structure feature of targets by classifying residues into three states (alpha helix, beta strand, and coil), as in DeepFusionDTA [29], realizing sequential representation of target structural features. Inspired by the above classification method, we divided residues into seven categories according to their physicochemical properties, as shown in Table 5. The target sequence is then transformed into a 7-category physicochemical property sequence that explicitly represents the physicochemical features of targets and significantly reduces the redundancy of protein features, shrinking the dimension of the feature matrix [30]. The above three target representations are then encoded to feature matrices $F_{seq} \in \mathbb{R}^{bs\times 1200}$, $F_{ss} \in \mathbb{R}^{bs\times 1200}$, and $F_{phyche} \in \mathbb{R}^{bs\times 1200}$ via label encoding.

**Table 4** Atomic feature set

| Atom features | Classification |
|---|---|
| Atomic number | 1, 6, 7, 8, 9, 15, 16, 17, 35, 53 |
| Atomic degree | 0, 1, 2, 3, 4, 5 |
| Hydrogen number | 0, 1, 2, 3, 4 |
| Implicit valence electron number | 0, 1, 2, 3, 4, 5 |
| Formal charge | −1, 0, 1 |
| Aromatic | 0, 1 |
| Hybrid | S-, SP-, SP2-, SP3-, SP3D-, SP3D2 |

**Table 5** Classification of targets' physical and chemical properties

| Physicochemical properties | Amino acid |
|---|---|
| Hydrophobic (nonpolar) | A, C, G, I, L, M, V, F, W |
| Hydrophilic (polar) | S, T, N, Q, P, U |
| Acidic (negatively charged) | D, E |
| Basic (positively charged) | K, R, H |
| Aromatic | F, W, Y |
| Sulfur-containing | C, M |
| Nonstandard | X, Z, O |

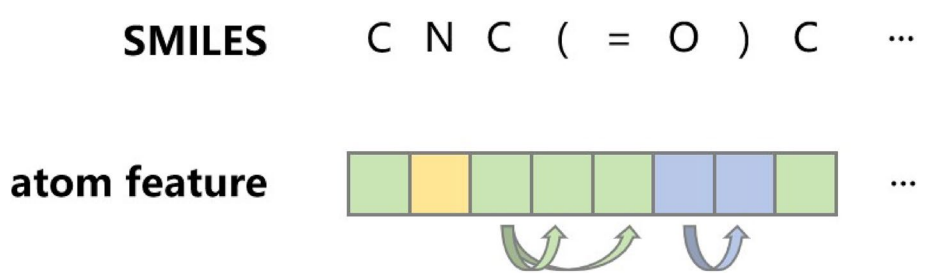



**Fig. 3** Encoding method for atom features. For nonatomic bits, such as "(," "=," and ")," the adjacency is simulated by sequentially filling in the features of the previous atom

## Proposed method

SMFF-DTA applies feature fusion and feature extraction in the Model Learning part after obtaining multifeature inputs of drugs and targets. In this section, we extract both local and global features simultaneously via a self-defined feature encoder. In addition, to capture crucial interaction features, we design a multiple attention block to focus closely on the interactions directly and indirectly. Since predicting binding affinity is a regression task, we obtain the final prediction value by using multiple fully connected layers in the last prediction section. Figure 4 shows the architecture of SMFF-DTA.

### *Input embedding*

The drug atom feature matrix $F_{atom}$ obtained via one-hot encoding is transformed into a dense feature matrix $X_{atom} \in \mathbb{R}^{bs\times 100\times D_{d1}}$ via a linear layer, where $D_{d1}$ represents the required dimensions. The high-dimensional sparse vectors are converted into low-dimensional dense vectors via linear transformation, improving the efficiency and compactness of feature representation. However, label encoding has certain limitations. It typically assumes that there is an order relationship between categories, which may affect our requirements for unordered classification. Therefore, we introduce an embedding layer before feature extraction and map the discrete inputs into continuous vector representations, converting the drug SMILES matrix $F_{smi}$ to $X_{smi} \in \mathbb{R}^{bs\times 100\times D_{d2}}$ and the target feature matrices $F_{seq}$, $F_{ss}$, and $F_{phyche}$ to $X_{seq} \in \mathbb{R}^{bs\times 100\times D_{t1}}$, $X_{ss} \in \mathbb{R}^{bs\times 100\times D_{td2}}$, and $X_{phyche} \in \mathbb{R}^{bs\times 100\times D_{t3}}$, respectively. Through the embedding layer, the model can better understand the input data and learn the complex relationships between different categories to improve model performance.

### *Feature encoder*

For sequences, features in local and global modes are complementary. Global features provide an overview of the target to capture global patterns and long-distance dependencies, whereas local features reflect specific functional regions. Therefore, we design a feature encoder to extract local features with a three-layer 1D CNN and extract global features with a BiGRU:

$$H_t^{(l+1)} = \mathrm{RELU}\left(\mathrm{CNN}_{1d}\left(W_t^{(l)}, b_t^{(l)}, H_t^{(l)}\right)\right), \quad l = 0, 1, 2 \tag{1}$$

$$H_t = \mathrm{BiGRU}(x_t, H_{t-1}) \tag{2}$$

where $H_t^{(l)}$ is the feature representation of layer l, $W_t^{(l)}$ and $b_t^{(l)}$ are learnable weights and offsets, $x_t$ is the input of time step t, and $H_t$ is the final merged hidden state of time step t. Then, adaptive average pooling is used to further address $H_t$ to preserve the most important feature through dimensionality reduction. To further enhance the representation capability of the model, we also incorporate the squeeze and excitation (SE, shown in Fig. 5) module [31] into the feature encoder. This helps us to explicitly model the interdependence between channels

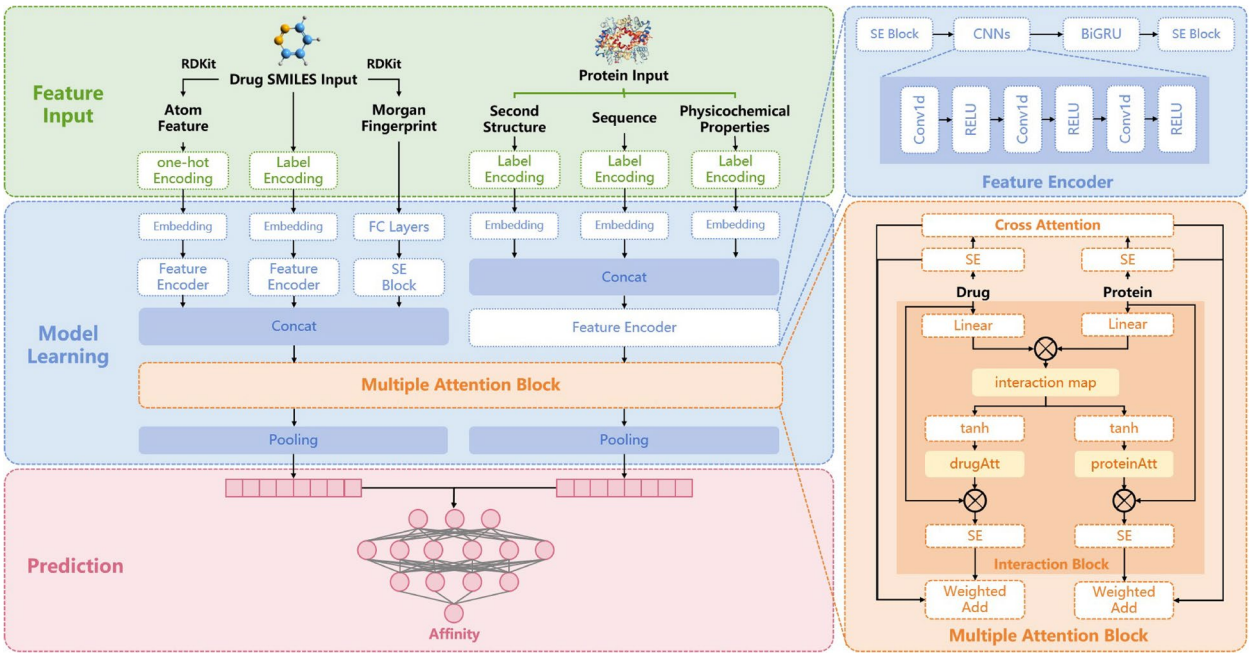



**Fig. 4** Architecture of SMFF-DTA. The overall pipeline is shown on the left side. The internal details of the feature encoder and the specific implementation process of the multiple attention block are shown on the right side

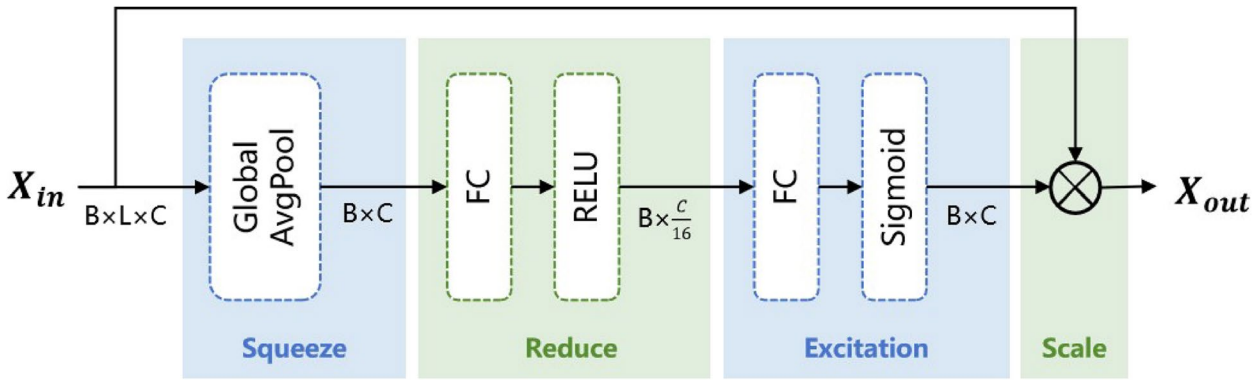



**Fig. 5** Details of the squeeze-and-excitation (SE) block. B represents the batch size, L represents the length, and C represents the channel dimension

and adaptively recalibrate the weights of different channels, realizing self-attention on channel dimensions [32]. We employ the SE block after the embedding layer and after feature extraction, allowing the network to focus better on more significant features, thereby indirectly reducing the impact of noise.

Furthermore, each type of feature in drug or target inputs is not independent; thus, we select efficient feature fusion strategies to further utilize the complementarity among multiple features, strengthening the model's robustness by lowering the uncertainty of each type of feature [33]. Since the tensor form of SMILES and atomic feature representation are different from that of the Morgan fingerprint, we choose the mid-fusion method to fuse drug features. In other words, we concatenate the features that are extracted separately, using a feature encoder to extract SMILES and atomic features while using an MLP to transform Morgan fingerprints into higher-level representations. In contrast to those of drugs, the tensor forms of target inputs are consistent. Therefore, we directly concatenate the inputs to extract the overall representation via the feature encoder to fully utilize the explicit information provided by the secondary structure and physicochemical properties though early fusion.

#### *Multiple attention block*

To better capture important drug–target interactions, we design a multiple attention block to extract interaction features in both indirect and direct ways. We improve the attention part of AttentionDTA [34], on the basis of which the interaction block is constructed. Interaction information is indirectly extracted by obtaining the interaction weight graph, and an SE block is added to reshape the channel weights. In addition, we capture interaction information directly through a multihead cross-attention mechanism, constructing explicit interactions between drugs and targets to take full advantage of their correlations [35], as shown in Fig. 6. Furthermore, parameter sharing is realized between drug attention and target attention, simplifying the computation to some extent. Taking drug attention as an example, we first recalibrate feature channels with the SE block and then map the input Query ($Q_d$), Key ($K_d$), and Value ($V_d$) into multiple subspaces via linear transformations. To realize information exchange between drugs and targets, a weighted sum of target values ($V_t$) is performed by calculating the attention score in each subspace. Ultimately, the new representation of the drug feature is generated by concatenating the outputs from each head together.

Extracted features acquired in direct and indirect ways are then weighted and fused. After maximum pooling, the most important feature is captured to represent the final feature:

$$W_{\text{score}} = \text{Softmax}(X_1 \cdot X_2) \tag{3}$$

$$X_{\text{new}} = X_1 \cdot W_{\text{score}} + X_1 + X_2 \cdot (1 - W_{\text{score}}) + X_2 + X_{\text{origin}} \tag{4}$$

$$X_{\text{out}} = \text{MaxPool}(X_{\text{new}}) \tag{5}$$

where $X_1$ and $X_2$ represent the features after indirect and direct attention, respectively, and where $X_origin$ represents the input features reshaped by the SE block.

#### *Prediction*

The final features of the drug and target obtained via concatenation are transmitted to the MLP with four fully connected layers, which predict the affinity value as the output:

$$X_{\text{in}} = \text{Concat}(X_d, X_t) \tag{6}$$

$$X_{\text{out}} = \text{LeakyReLU}(\text{FC}(X_{\text{in}})) \tag{7}$$

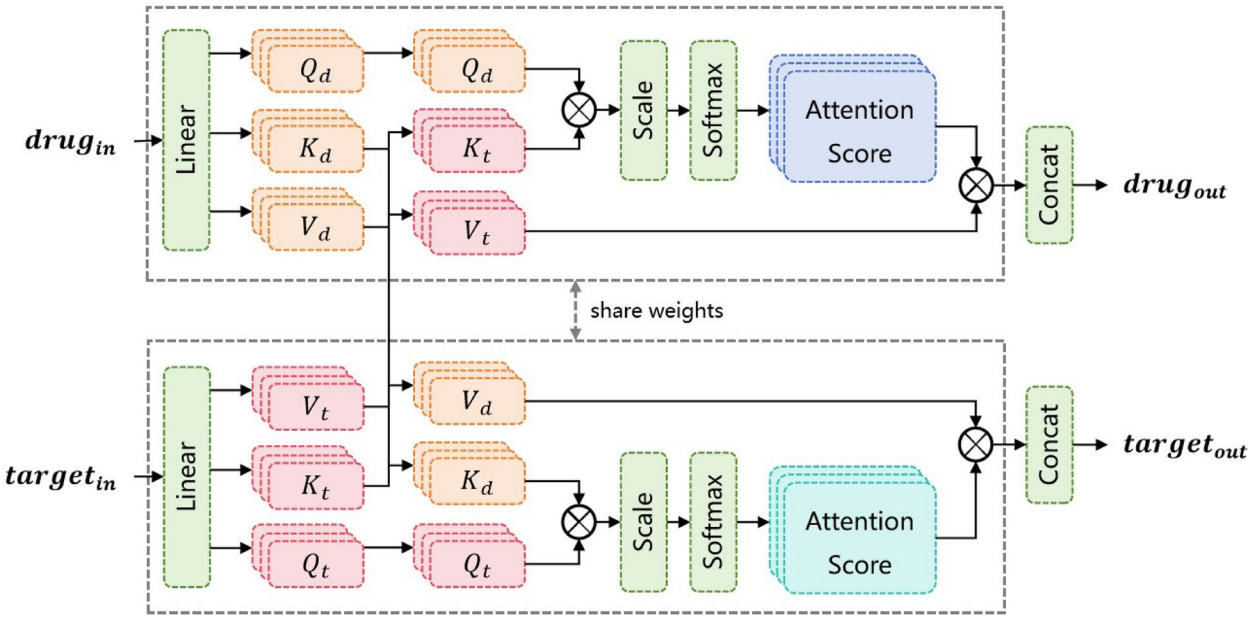



**Fig. 6** Details of multihead cross-attention between the drug and target

### Model training

Our model is trained via 5-fold CV, randomly dividing the dataset into six parts, one of which is utilized as an independent test set, while the remaining five parts are trained and verified via nested cross-validation. MSE, as the loss function for model training, assesses the gap between the predicted and actual values:

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{8}$$

where $p_i$ and $y_i$ are the predicted affinity and actual value, respectively. The AdamW optimizer is used for parameter training. Model performance is monitored according to the MSE values, preventing the model from overfitting via an early stop strategy. The hyperparameter settings of our model are shown in Table 6. The batch size and learning rate on KIBA are increased in proportion to the size of the dataset, which is approximately four times larger than Davis's. Additionally, the results of the parameter experiments on the batch size and learning rate are shown in Additional file 1: Table 1, and the results of the experiments on the number of attention heads are shown in Additional file 1: Table 2 and Table 3.

### Evaluation metrics

Regarding affinity prediction as a regression task, we evaluate model performance in terms of the mean squared error (MSE), $R_m^2$ and concordance index (CI). The MSE measures the deviation between the predicted and actual values and assesses the error of model prediction:

**Table 6** Hyper-parameters settings of SMFF-DTA

| Hyper-parameters | Settings |
|---|---|
| $D_{d1}$, $D_{d2}$ | 64, 128 |
| $D_{t1}$, $D_{t2}$, $D_{t3}$ | 64, 6, 15 |
| Learning rate | [1e-6, **5e-6**, 8e-6, 1e-5] (Davis), 2e-5(KIBA) |
| Batch size | [**32**, 64, 128, 256](Davis), 128(KIBA) |
| Interaction Block Head Number | [2, **4**, 8, 16] |
| Cross Attention Head Number | [2, 4, **8**, 16] |

The best parameters selected according to the experimental results are shown in bold

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{9}$$

where $p_i$ and $y_i$ are the predicted affinity and actual value, respectively. $R_m^2$ [36] reflects the external predictive performance of a model, which is widely used to verify the regression-based quantitative structure–activity relationship (QSAR) model and evaluate the correlation:

$$R_m^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{10}$$

where $r^2$ and $r_0^2$ are the square correlation coefficients with and without intercepts, respectively. Generally, the predictive performance of a model whose $R_m^2$ exceeds 0.5

is considered acceptable. CI [37] measures whether the predicted values of two random drug–target pairs are in the same order as their true values, assessing the accuracy of the model:

$$CI = \frac{1}{Z} \sum_{y_i > y_j} h(p_i - p_j) \tag{11}$$

$$h(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases} \tag{12}$$

where $Z$ is a normalized constant and $h(x)$ is step function.

**Abbreviations**

| | |
|---|---|
| DTA | Drug-target binding affinity |
| SMILES | Simplified molecular linear input specification |
| CNN | Convolutional neural network |
| RNN | Recurrent neural network |
| GAN | Generative adversarial network |
| GCN | Graph convolutional network |
| GAT | Graph attention network |
| MSE | Mean square error |
| CI | Consistency index |
| Kd | Kinase dissociation constant |
| Ki | Inhibition constant |
| IC50 | Half maximum inhibitory concentration |
| 1D CNN | 1D convolutional neural network |
| BiGRU | Bidirectional gated recurrent unit |
| MLP | Multilayer perceptron |
| CV | Cross-validation |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02222-x.

Additional file 1: Tables 1-3. The document shows the experimental results of some parameters of SMFF-DTA. Table 1 - The parameter experiments on batchsize and learning rate in Davis. Table 2 - The experiments on the number of attention heads in Cross-Attention. Table 3 - The experiments on the number of attention heads in Interaction Block.

**Acknowledgements**
Not applicable.

**Authors' contributions**
Zhijun Xia contributed to the initial draft and the design and implementation of the experiments. Runqiu Feng and Tongyu Han were responsible for data collection. Wenqian Yu and Hanyu Wang made reference preparation. Xun Wang and Xingguang Wang provided experimental guidance and revised the manuscript. All authors read and approved the final manuscript.

**Funding**
This work was supported by the National Natural Science Foundation of China [No.61972416], Natural Science Foundation of Shandong Province [No. ZR2022LZH009], GHfund C (202407035455), National Key R&D Program of China [No.2021YFA1000103-3].

**Data availability**
The datasets and code supporting the conclusions of this article are available in the Zenodo repository [https://doi.org/10.5281/zenodo.15054908]. All data generated or analysed during this study are included in this published article, its supplementary information files and publicly available repositories.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

Received: 12 February 2025 Accepted: 24 April 2025
Published online: 09 May 2025

**References**

1. Berdigaliyev N, Aljofan M. An overview of drug discovery and development. Futur Med Chem. 2020;12(10):939–47.
2. Zhu X, Liu J, Zhang J, Yang Z, Yang F, Zhang X. FingerDTA: a fingerprint-embedding framework for drug-target binding affinity prediction. Big Data Min Anal. 2022;6(1):1–10.
3. Kimber TB, Chen Y, Volkamer A. Deep learning in virtual screening: recent applications and developments. Int J Mol Sci. 2021;22(9):4435.
4. Zeng X, Li SJ, Lv SQ, Wen ML, Li Y. A comprehensive review of the recent advances on predicting drug-target affinity based on deep learning. Front Pharmacol. 2024;15:1375522.
5. Jiang M, Shao Y, Zhang Y, Zhou W, Pang S. A deep learning method for drug-target affinity prediction based on sequence interaction information mining. PeerJ. 2023;11:e16625.
6. Wang H. Prediction of protein–ligand binding affinity via deep learning models. Brief Bioinform. 2024;25(2):bbae081.
7. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug-target interaction predictions. Brief Bioinform. 2015;16(2):325–37.
8. He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines. J Cheminformatics. 2017;9:1–14.
9. Chauhan NK, Singh KA, review on conventional machine learning vs deep learning. In: 2018 International conference on computing, power and communication technologies (GUCON). IEEE; 2018. pp. 347–52.
10. Askr H, Elgeldawi E, Aboul Ella H, Elshaier YA, Gomaa MM, Hassanien AE. Deep learning in drug discovery: an integrative review and future challenges. Artif Intell Rev. 2023;56(7):5975–6037.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
12. Zaremba W. Recurrent neural network regularization. 2014. arXiv preprint arXiv:1409.2329.
13. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998–6008. https://doi.org/10.48550/arXiv.1706.03762.
14. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018;34(17):i821–9.
15. Zhao L, Wang J, Pang L, Liu Y, Zhang J. GANsDTA: Predicting drug-target binding affinity using GANs. Front Genet. 2020;10:1243.
16. Li W, Zhou Y, Tang X. TF-DTA: A Deep Learning Approach Using Transformer Encoder to Predict Drug-Target Binding Affinity. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2023. pp. 418–421.
17. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw. 2008;20(1):61–80.
18. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. 2017. arXiv preprint arXiv:1710.10903.
19. Tang W, Zhao Q, Wang J. LLMDTA: Improving Cold-Start Prediction in Drug-Target Affinity with Biological LLM. In: International Symposium on Bioinformatics Research and Applications. Springer; 2024. pp. 152–163.

20. Deng J, Zhang Y, Pan Y, et al. Multidta: drug-target binding affinity prediction via representation learning and graph convolutional neural networks. Int J Mach Learn Cyber. 2024;15:1–10. https://doi.org/10.1007/s13042-023-02042-x.
21. Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. J Med Chem. 2005;48(12):4111–9.
22. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011;29(11):1046–51.
23. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J Chem Inf Model. 2014;54(3):735–43.
24. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci. 1988;28(1):31–6.
25. Yang X, Niu Z, Liu Y, Song B, Lu W, Zeng L, et al. Modality-DTA: multimodality fusion strategy for drug-target affinity prediction. IEEE/ACM Trans Comput Biol Bioinforma. 2022;20(2):1200–10.
26. Landrum G, et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. Greg Landrum. 2013;8(31.10):5281.
27. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742–54.
28. Deng L, Zeng Y, Liu H, Liu Z, Liu X. DeepMHADTA: prediction of drug-target binding affinity using multi-head self-attention and convolutional neural network. Curr Issues Mol Biol. 2022;44(5):2287–99.
29. Pu Y, Li J, Tang J, Guo F. DeepFusionDTA: drug-target binding affinity prediction with information fusion and hybrid deep-learning ensemble model. IEEE/ACM Trans Comput Biol Bioinforma. 2021;19(5):2760–9.
30. Che C, Zhu M, Zhu Y, Zhang Q, Zhou D, Wang B. A protein embedding model for drug molecular screening. In: 2020 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE; 2020. pp. 251–254.
31. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE; 2018. pp. 7132–7141.
32. Kyro GW, Brent RI, Batista VS. Hac-net: A hybrid attention-based convolutional neural network for highly accurate protein-ligand binding affinity prediction. J Chem Inf Model. 2023;63(7):1947–60.
33. Jones D, Kim H, Zhang X, Zemla A, Stevenson G, Bennett WD, et al. Improved protein-ligand binding affinity prediction with structure-based deep fusion inference. J Chem Inf Model. 2021;61(4):1583–92.
34. Zhao Q, Duan G, Yang M, Cheng Z, Li Y, Wang J. AttentionDTA: drug-target binding affinity prediction by sequence-based deep learning with attention mechanism. IEEE/ACM Trans Comput Biol Bioinforma. 2022;20(2):852–63.
35. Jin Z, Wu T, Chen T, Pan D, Wang X, Xie J, et al. CAPLA: improved prediction of protein–ligand binding affinity by a deep learning approach based on a cross-attention mechanism. Bioinformatics. 2023;39(2):btad049.
36. Ojha PK, Mitra I, Das RN, Roy K. Further exploring rm2 metrics for validation of QSPR models. Chemometr Intell Lab Syst. 2011;107(1):194–205.
37. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika. 2005;92(4):965–70.


## Publisher's Note