Methods 207 (2022) 103–109

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

MultiscaleDTA: A multiscale-based method with a self-attention mechanism for drug-target binding affinity prediction

Haoyang Chen [a,b,1], Dahe Li [c,1], Jiaqi Liao [b], Lesong Wei [d,*], Leyi Wei [a,b,*]

[a] School of Mathematics and Statistics, Hainan Normal University, Hainan, China
[b] School of Software, Shandong University, Jinan, China
[c] Beidahuang Industry Group General Hospital, Harbin 150001, China
[d] Department of Computer Science, University of Tsukuba, Tsukuba 3058577, Japan

ABSTRACT

The task of predicting drug-target affinity (DTA) plays an increasingly important role in the early stage of *in silico* drug discovery and development. Currently, a variety of machine learning-based methods have been presented for DTA prediction and achieved outstanding performance, which is beneficial for speeding up the development of new drugs. However, most convolutional neural networks (CNNs) based methods ignore the significance of information from CNN layers with different scales for DTA prediction. In addition, each feature provides different contributions to the final task. Therefore, in this study, we propose a novel end-to-end deep learning-based framework, MultiscaleDTA, to predict drug-target binding affinity. MultiscaleDTA incorporates multi-scale CNNs and a self-attention mechanism to capture multi-scale and comprehensive features for characterizing the intrinsic properties of drugs and targets. Extensive experimental results on both regression and binary classification tasks demonstrate that MultiscaleDTA achieves competitive performance compared to state-of-the-art methods.

1. Introduction

Identifying compounds that can generate the reaction with the target protein and then control the progression of diseases is the aim of drug discovery. The pipeline of new drug discovery and development is cumbersome and complex and relies on many factors. Accurately predicting the binding affinity between the drug-target pair is one of the most important steps in the early stage of drug discovery and design and is beneficial to decipher potential mechanisms of drug off-target adverse and action events [1]. Experimentally determining the drug-target binding affinity remains the most effective method. However, it is time-consuming and prohibitively expensive, especially considering that there are over 5000 possible protein targets [2] and more than 100 million potential drug compounds [3]. Therefore, it is necessary to develop computational methods to rapidly predict the binding affinity to facilitate the process of screening drug candidate molecules, narrowing the search space of wet experiments [4,5].

The existing computational methods for DTA prediction mainly fall into two categories, including structure-based methods and structure-free methods. Structure-based methods, such as molecular docking, utilize the known three-dimensional structures of drug and protein molecules to measure the binding affinity between them. Although these methods can accurately measure the binding affinity, they suffer from two main limitations: (1) the scarcity of high-quality manually annotated 3D structures of proteins and obtaining such structures remains challenging, and (2) requiring high computational resources.

To mitigate the limitations, a number of structure-free methods (*i.e.*, machine learning-based methods and deep learning-based methods) for DTA prediction have been proposed to speed up potential drug screening. Such methods generally consist of two steps: feature extraction and regression procedure. First, the features of drug compounds and target proteins are extracted from primary sequences. Second, the obtained features are used as inputs of regressors or networks to measure specific values. For example, Ballester *et al.* proposed a random forest-based score function to predict the protein–ligand binding affinity by taking the occurrence times of specific protein–ligand atom type pairs interacting within a certain distance as features [6]. SimBoost uses the drug and target similarities and a matrix factorization model to extract the effective binding features for predicting DTA based on a gradient boosting tree model [7]. Betsabeh *et al.* used the multi-source information from different similarity measures and the K-nearest neighbor algorithm to build a DTA prediction model based on the gradient boosting machine [8]. Fergus *et al.* demonstrated that combining a diverse group of ligand-based features in machine learning scoring functions can

[*] Corresponding authors at: School of Mathematics and Statistics, Hainan Normal University, Hainan, China (L Wei).
E-mail addresses: wei.lesong.ww@alumni.tsukuba.ac.jp (L. Wei), weileyi@sdu.edu.cn (L. Wei).
[1] These authors contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.ymeth.2022.09.006
Received 13 August 2022; Received in revised form 15 September 2022; Accepted 19 September 2022
Available online 23 September 2022

improve their performance for DTA prediction [9].

These methods have the ability to generalize to drug-target pairs with no similarity to any ones in the training set and are high sequence sensitively, which means they can output distinct binding affinities for highly similar drug-target pairs. However, they are limited by the expressiveness of the handcrafted features that rely on prior knowledge, leading to the limited performance of models. Recently, with the remarkably successful application in various fields of deep learning, deep learning-based methods for DTA prediction have gained unprecedented attention, which can use an end-to-end way to automatically extract feature representations of the drug and target and achieve superior performance [10–17]. For example, Hakime *et al.* proposed DeepDTA in 2018, which used the convolutional neural networks (CNNs) to extract features from drug and protein sequences separately to predict the drug-target binding affinity [14]. In 2019, based on DeepDTA, Hakime *et al.* proposed another model WideDTA, which not only used ligand SMILES and protein sequences as inputs but also adopted ligand maximum common substructures and protein domains and motifs to extract effective features to represent the drug-target pair [12]. MATT_DTI builds a relation-aware self-attention block to learn the correlations between atoms for the drug representation and then uses the multi-head attention mechanism to combine the drug and protein representations for final prediction [11]. To extract the similarity between drug representation and protein representation, DeepCDA employs a hybrid network consisting of CNN and long-short-term memory (LSTM) to capture the features of drugs and targets, respectively, and then proposes a two-sided attention mechanism to learn the binding representation [10]. To take advantage of the important topological structure information hidden in these molecules, several models have been built to model two-dimensional graph structure information of drugs and targets. For example, GraphDTA represents drugs graphically and extracts compound representations using graph neural networks (GNN) to predict drug-target affinity [18]. Tian *et al.* improved GraphDTA model from a two-channel model to a three-channel model, interpreted the target/protein sequence as time series, and used LSTM network to extract its features [19]. DGraphDTA uses contact maps as the inputs of proteins based on GraphDTA, and uses GNNs to extract drug and protein structural information, respectively [20].

Although these deep learning-based methods have achieved satisfactory performance on the DTA prediction task, there still exist some problems. First, most of them ignore the heterogeneous information from different neural network layers, which tends to lose parts of task-related information. In addition, they also do not consider the fact that the features from the neural network may contain some redundant elements, which can interfere with the final task, leading to relatively poor performance.

To overcome the above-mentioned problems, we propose a novel end-to-end deep learning-based method, called MultiscaleDTA, for DTA prediction. To be specific, we first design a multi-scale CNNs feature extractor to capture multi-scale local information from primary drug (target) sequences. Subsequently, a self-attention mechanism is adopted to measure the contribution of each feature to the final DTA prediction for obtaining a more effective multi-scale feature representation. Next, the feature representations from different CNN layers are concatenated as the final drug (target) representation. MultiscaleDTA further concatenates the drug and target representations and sends the combined features into fully connected layers to predict DTA. The effective and multi-scale supervision information can help MultiscaleDTA to predict drug-target affinity accurately only based on primary sequence information. Through extensive experiments on benchmark datasets, we demonstrate that compared to other state-of-the-art methods, MultiscaleDTA not only achieves superior performance on DTA prediction, but also obtains competitive performance on drug-target interaction identification.

2. Material and methods

2.1. Datasets

In this work, drug-target binding affinity prediction is formulated as a regression task, and we select two public benchmark datasets proposed for DTA prediction to evaluate the performance of the proposed MultiscaleDTA, including Davis [21] and KIBA [22] datasets. The Davis dataset contains 30,056 drug-target pairs, in which the number of unique drugs is 68 and the number of unique targets is 442. The kinase dissociation constant K_d is taken to measure the binding affinity. The higher K_d value means the binding between the drug-target pair is lower. Following the previous study [23–26], the K_d values are transformed into log space as follows:

$$pK_d = -\log 10\left(\frac{K_d}{10^9}\right) \tag{1}$$

The KIBA dataset consists of 229 unique drugs and 2111 unique targets, which form 118,254 different drug-target pairs. Differing from the Davis dataset, it adopts KIBA scores as binding affinities, which incorporate different information sources:K_d, K_i (inhibition constant), and IC50(the half-maximal inhibitory). In these two datasets, the SMILES strings of drugs are collected from PubChem compound database [27] based on PubChem CIDs and the target's protein sequences are collected from UniProt database [28].

In addition, we also assess the performance of MultiscaleDTA in drug-target interaction (DTI) prediction, which is referred to as a binary classification task. Two publicly available datasets are taken as benchmark datasets: Human and Caenorhabditis elegans(*C.elegans*) datasets. The positive DTI pairs in them are derived from Matador [29] and DrugBank4.1 [30] and the highly credible negative pairs are generated by using a systematic screening framework. The details of these four datasets are reported in Table 1.

2.2. Model architecture

Fig. 1 shows the overall framework of the proposed MultiscaleDTA. The MultiscaleDTA model mainly consists of CNN networks and the self-attention mechanism. For the drug SMILES (Simplified Molecular Input Line Entry System) sequence and the protein amino acid sequence, we first employ the one-hot technique to transform them into numerical data, which meet the requirement of the neural network. Next, the obtained numerical data are sent into two three-layer multi-scale CNNs, respectively, to extract local and multi-scale information. Then, the features from each CNN layer are multiplied by a self-attention matrix to learn more effective features. Subsequently, the optimized features from different CNN layers are concatenated as the representation of the drug (target) molecule, which contain multi-scale and more comprehensive information to characterize the intrinsic properties of molecules. Finally, we concatenate the representations of the drug and target and send the combined features into fully connected layers to conduct DTA prediction.

2.3. Input representation

For the drug molecule, the SMILES is adopted to represent it, which is a chemical notation that uses short ASCII strings to describe the chem-

Table 1.S
ummary of the benchmark datasets.

Datasets	Task type	Proteins	Compounds	Interactions
Davis	Regression	442	68	30,056
KIBA	Regression	229	2111	118,254
Human	Classification	2001	2726	6728
C. elegans	Classification	1876	1767	7786

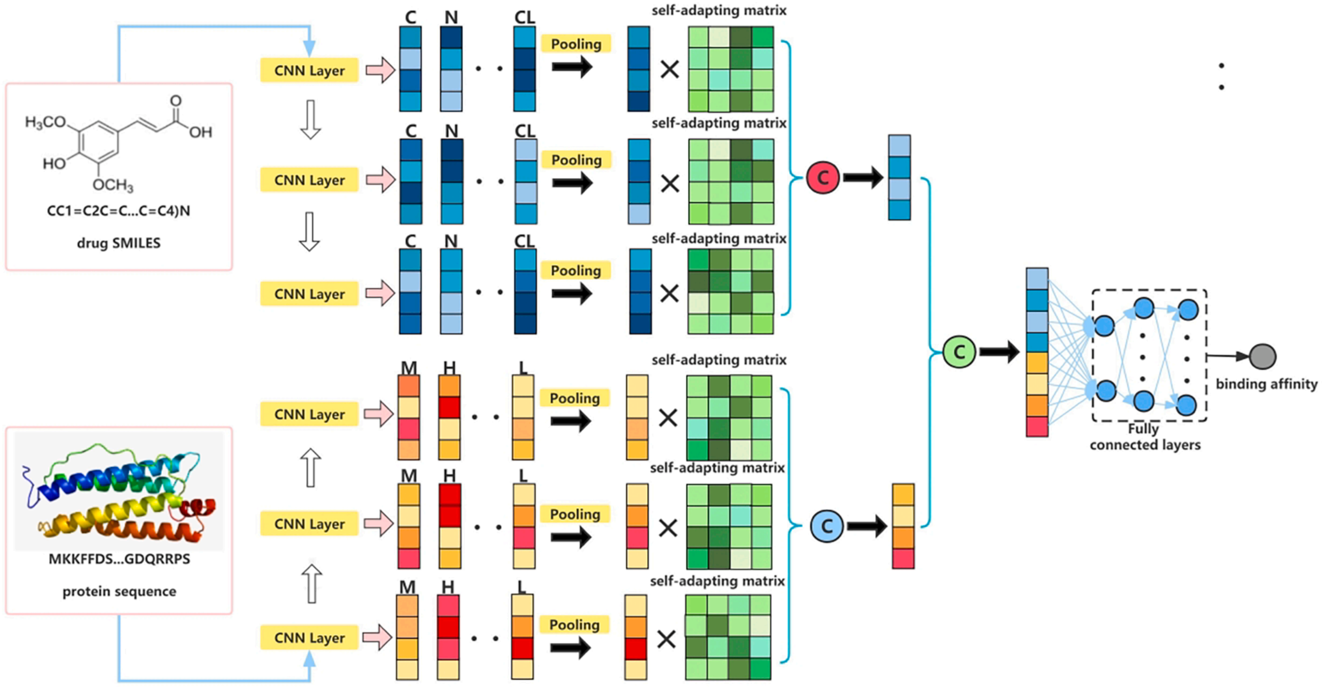

Fig. 1. The overall architecture of the proposed MultiscaleDTA. Firstly, the drug sequence and protein sequence are embedded by the one-hot technique. Then, the two three-layer multi-scale CNNs with the self-attention mechanism are designed to extract multi-scale and comprehensive features of the drug and target molecules. Finally, the features of the drug and target molecules are concatenated, and the combined features are imported into the fully connected layers to predict the binding affinity.

ical species structure. We use one-hot technique to embed it. Specifically, a vocabulary is constructed to denote each character in SMILES [31]. Then each character in SMILES is projected with an integer between 1 and 64. For each drug SMILES string D, we can get a numerical vector as follows:

$$D = [s_1, s_2, \cdots, s_n], \tag{2}$$

where n is the length of the drug SMILES string and each element denotes an integer representing the character type.

For the target molecule, it is represented by an amino acid string. Similar to the drug molecule, we also construct a vocabulary, in which each integer between 1 and 20 represents a type of the standard amino acid. For given a target sequence T of length l, a numerical vector is generated as follows:

$$T = [a_1, a_2, \cdots, a_l]. \tag{3}$$

Note that we set the largest length of SMILES to 100. If the length of a drug string is longer than 100, the excess parts will be truncated. And if the length is less than 100, the zero-padding technique will be adopted. For the target protein sequence, the maximum length is set to 1200. To meet the requirement of neural networks, they are processed like the drug.

2.4. CNN

The convolutional neural network is a classical type of feedforward neural network, which is commonly used to extract latent contextual semantic information of sequence-based data (*e.g.*, text) based on multiple convolution kernels [32]. Previous studies have illustrated that CNNs can effectively capture discriminative local features from primary protein sequences by using multiple types of convolution kernels, leading to satisfactory performance [33–37]. Inspired by these studies, we use multi-scale CNNs to code protein sequences into fixed-size feature vectors that contain multi-scale local information. Our CNN module consists of embedding, convolutional, activation and pooling layers. The embedding layer is used to transform the input vector into the matrix data. For the convolutional layers, they have three different sizes of convolutional kernels, including 96 × 96, 64 × 64, and 32 × 32. The output of each CNN layer is the obtained features and is further processed by the activation layer that is equipped with rectified linear unit (*ReLU*) and the pooling layer that is equipped with the max pooling function. To avoid over-fitting, the dropout function is also adopted. Specifically, for the drug SMILES string $D = [s_1, s_2, \cdots, s_n]$, we first send it into the embedding layer as follows:

$$M_D = Embedding(D), (4)$$

Then the obtained M_D is sent into CNN, activation, and pooling layers:

$$\begin{aligned}
F_D^{(1)} &= maxpooling\left(ReLU\left(CNN^{(1)}(M_D)\right)\right) \\
F_D^{(2)} &= maxpooling\left(ReLU\left(CNN^{(2)}\left(ReLU\left(CNN^{(1)}(M_D)\right)\right)\right)\right) \\
F_D^{(3)} &= maxpooling\left(ReLU\left(CNN^{(3)}\left(ReLU\left(CNN^{(2)}\left(ReLU\left(CNN^{(1)}(M_D)\right)\right)\right)\right)\right)\right)
\end{aligned} \tag{5}$$

For the target amino acid sequence $T = [a_1, a_2, \cdots, a_l]$, similar to the drug, we can get the following feature vectors:

$$\begin{aligned}
F_T^{(1)} &= maxpooling\left(ReLU\left(CNN^{(1)}(M_T)\right)\right) \\
F_T^{(2)} &= maxpooling\left(ReLU\left(CNN^{(2)}\left(ReLU\left(CNN^{(1)}(M_T)\right)\right)\right)\right) \\
F_T^{(3)} &= maxpooling\left(ReLU\left(CNN^{(3)}\left(ReLU\left(CNN^{(2)}\left(ReLU\left(CNN^{(1)}(M_T)\right)\right)\right)\right)\right)\right)
\end{aligned} \tag{6}$$

2.5. *The self-attention mechanism*

Each feature vector from different CNN layers may contain some redundant and task-irrelated elements, which will have a negative impact on the final DTA prediction. And each element in the feature vector has different contributions to final prediction. Therefore, inspired by the graph attention neural network (GAT) and borrow the idea of the self-attention mechanism to make the model focus on important parts in a feature vector, which provide more contribution to the DTA prediction. To be specific, we set up a learnable parameter matrix to assign weights to the vectors from different CNN layers and then multiply the matrix with the corresponding vector. Finally, the weighted vectors are concatenated as the final representation of the molecule as follows:

$$V = W_1F^{(1)}\|W_2F^{(2)}\|W_3F^{(3)}, \tag{7}$$

where W_i is the i-th learnable parameter matrix in i-th layer and $F^{(i)}$ represents the output vector from the i-th max pooling layer, and '||' denotes the concatenate operation.

For a given drug, we can get a vector $V_D = W_1F_D^{(1)}\Big\|W_2F_D^{(2)}\Big\|W_3F_D^{(3)}$. Similarly, for a target protein sequence, we get $V_T = W_1F_T^{(1)}\Big\|W_2F_T^{(2)}\Big\|W_3F_T^{(3)}$.

2.6. Prediction

After obtaining the representations of the drug V_D and target V_T, we further concatenate them together as the representation of a drug-target pair, that is $V_{final} = V_D\|V_T$. Then, we feed it into a three-layer fully connected network with the *ReLU* activation function to predict the binding affinity.

$$y = FC\left(W^{(1)}, W^{(2)}, W^{(3)}, V_{final}\right) \tag{8}$$

where $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$ are the weight parameters of the fully connected layers.

For a group of drug-target pairs in the training dataset and corresponding truth affinity values, mean squared error (MSE) is taken as the loss function to train the various weight parameters in our model as follows:

$$\text{loss} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{9}$$

Where y_i and $\widehat{y}_i$ are the predictive affinity and the truth affinity of i-th drug–target pair, and n is the number of the drug–target pairs. In addition, when we train our model to prediction DTI, the loss function MSE is replaced with the cross-entropy function, which is the commonly used loss function for classification.

2.7. Hyperparameter settings

Training a DTA model requires hyperparameter settings, and there are many hyperparameters in MultiscaleDTA. Since it takes hours to train a model, some parameters are set based on the human experience. Our hyperparameter settings are shown in Table 2.

Table 2
The hyperparameter settings using human experience.

Hyperparameter	Setting
Epoch	2000
Learning rate	0.0008
Batch size	512
Optimizer	Adam
Dropout	0.2
Pooling method	max pooling
Fully connected layers after concatenation	3

3. Result and discussion

3.1. Performance evaluation metrics

Concordance index (CI), a model evaluation index proposed by GÖnen and Heller [38], is designed to calculate the difference between the predicted value of the model and the ground truth. CI is defined as follows:

$$\text{CI} = \frac{1}{Z}\sum_{d_y > d_x} h\left(b_x - b_y\right) \tag{10}$$

where b_x is the prediction value for d_x, b_y is the prediction value for d_y, *h(x)* is the step function and Z is the normalized hyperparameter. Commonly, the step function h(x) is defined as follows:

$$h(x) = \begin{cases} 1, ifx > 0 \\ 0.5, ifx = 0 \\ 0, ifx < 0 \end{cases} \tag{11}$$

Regression toward the mean (r_m^2 index) is a measure evaluating the external predictive performance of a model. If a variable is very large, then r_m^2 means how much it tends to approach the average next time. r_m^2 index is calculated as follows:

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{12}$$

where r is the squared correlation coefficients with intercepts and r_0 is the coefficients without intercepts.

3.2. Compare with DTI prediction models in classification tasks

We also evaluate our model's performance in the drug-target interaction prediction (DTI), which belongs to the binary classification task to identify whether a given drug-target is interactive or not. For a fair comparison, we conduct experiments on two benchmark datasets: Human dataset and *C.elegans* dataset and choose the area under the receiver operating characteristic curve (AUC), Precision, and recall as the performance evaluation metrics. We compare our model with traditional machine learning-based methods, including K-nearest neighbors (KNN), Random Forest (RF), L2-logistic (L2), and support vector machine (SVM), and several deep learning-based methods, including CPI-GNN [39], DrugVQA [40], and TransformerCPI [41]. Note that for a fair comparison, the input of DrugVQA is the structure-based features of proteins, while its another version VQA-SEQ that only takes primary amino acid information as input is displayed here. We implement all experiments five times with different seeds and report the average and standard deviation of results as the final result in DTI prediction. The best results of all methods are marked in bold.

The comparative results of MultiscaleDTA and competing methods on the Human dataset are presented in Table 3. It can be seen that except for SVM, MultiscaleDTA has better performance than other methods in terms of the best AUC, Precision, and Recall, which are 0.981, 0.947,

Table 3
The performance of MultiscaleDTA and baseline models on the Human dataset.

Model	AUC	Precision	Recall
KNN	0.860	0.927	0.798
RF	0.940	0.897	0.861
L2	0.911	0.913	0.867
SVM	0.910	**0.966**	**0.969**
GraphDTA	0.960(±0.005)	0.882(±0.040)	0.912(±0.040)
GCN	0.956(±0.004)	0.862(±0.006)	0.928(±0.010)
CPI-GNN	0.970	0.918	0.923
DrugVQA(VQA-seq)	0.964(±0.005)	0.897(±0.004)	0.948(±0.003)
TransformerCPI	0.973(±0.002)	0.916(±0.006)	0.925(±0.006)
MultiscaleDTA	**0.981(±0.004)**	0.947(±0.005)	0.942(±0.006)

and 0.942, respectively. For the method SVM, although it achieves the best Precision and Recall on Human dataset that is a relatively small dataset, when it runs on a larger dataset (*i.e.*, *C.elegans* dataset), its performance has a great drop. The reason may be that traditional machine learning-based methods need to export knowledge to characterize the intrinsic properties of molecules, leading to poor generalization.

Table 4 summarizes the results of MultiscaleDTA and other methods on *C.elegans* dataset. As shown in Table 4, we can see that MultiscaleDTA significantly outperforms other methods in terms of AUC, Precision, and Recall. To be specific, when compared to the runner-up TransformerCPI, MultiscaleDTA improves AUC from 0.988 to 0.994 (a relative improvement of 0.6 %), Precision from 0.952 to 0.982 (a relative improvement of 3.2 %), and Recall from 0.953 to 0.971 (a relative improvement of 1.9 %), which indicates the superiority of our model in the DTI prediction task.

3.3. Compare with DTA prediction models in regression tasks

To evaluate the performance of the proposed MultiscaleDTA in DTA prediction, we compare it with several state-of-the-art deep learning-based methods on Davis and KIBA datasets, including DeepDTA, WideDTA, MT-DTI, DeepCDA, MATT-DTI, and GraphDTA. Note that DGraphDTA [24] is not adopted since it uses the structure-based features to represent proteins. Two feature-based methods also are selected for comparison, including KronRLS and SimBoost. The best results of all methods are marked in bold font.

In Table 5, the comparative results between MultiscaleDTA and competing methods on the Davis dataset are listed. As shown in Table 5, we can see that when compared to feature-based methods, MultiscaleDTA improves CI index by 0.027–0.026 and r_m^2 index by 0.094–0.331 and reduces MSE by 0.082 and 0.179. In addition, when compared to deep learning-based methods, MultiscaleDTA also significantly outperforms them in terms of the best CI index, MSE and r_m^2 index, which are 0.898, 0.200, and 0.738, respectively.

The performance of MultiscaleDTA and competing methods on the KIBA dataset is summarized in Table 6. From this table, we can observe that compared to either feature-based methods or deep learning-based methods, MultiscaleDTA greatly improves the predictive performance. Each metric for evaluation, including CI index, MSE, and r_m^2 index, has achieved significant improvement. For CI and r_m^2 indexes, MultiscaleDTA can reach 0.893 and 0.793, which are 0.002–0.111 and 0.037–0.451 higher than competing methods.

The above discussed results illustrate the effectiveness and robustness of our model. There are two main reasons. Firstly, when compared to feature-based methods which highly rely on the expert knowledge-based handcrafted features as the input, MultiscaleDTA does not require the prior knowledge and can automatically extract effective and discriminative features from original data. Secondly, when compared to deep learning-based methods, MultiscaleDTA introduces a self-attention mechanism to make the model focus on important parts of each feature representation from different CNN layers. In addition, it constructs an informative feature profile by integrating the multi-scale information from different CNN layers, leading to a more comprehensive feature representation, which plays a crucial role in building a powerful predictor. Therefore, there is no surprise for the superiority of our model.

Table 4
The performance of MultiscaleDTA and baseline models on the *C. elegans* dataset.

Model	AUC	Precision	Recall
KNN	0.858	0.801	0.827
RF	0.902	0.821	0.844
L2	0.892	0.890	0.877
SVM	0.894	0.785	0.818
GraphDTA	0.974(±0.004)	0.927(±0.015)	0.912(±0.023)
GCN	0.975(±0.004)	0.921(±0.008)	0.927(±0.006)
CPI-GNN	0.978	0.938	0.929
TransformerCPI	0.988(±0.002)	0.952(±0.006)	0.953(±0.005)
MultiscaleDTA	**0.994(±0.003)**	**0.982(±0.009)**	**0.971(±0.005)**

Table 5
The performance of MultiscaleDTA and baseline models on the Davis dataset.

Model	CI(std)	MSE	r_m^2(std)
KronRLS	0.871 (±0.001)	0.379	0.407 (±0.005)
SimBoost	0.872 (±0.002)	0.282	0.644 (±0.006)
DeepDTA	0.878 (±0.004)	0.261	0.630 (±0.017)
WideDTA	0.886 (±0.003)	0.262	–
MT-DTI	0.887 (±0.003)	0.245	0.665 (±0.014)
DeepCDA	0.891 (±0.003)	0.248	0.649 (±0.009)
MATT-DTI	0.891 (±0.002)	0.227	0.683 (±0.017)
GraphDTA	0.893 (±0.001)	0.229	–
MultiscaleDTA	**0.898(±0.004)**	**0.200**	**0.738(±0.012)**

Table 6
The performance of MultiscaleDTA and baseline models on the KIBA dataset.

Model	CI(std)	MSE	r_m^2(std)
KronRLS	0.782 (±0.001)	0.441	0.342 (±0.001)
SimBoost	0.836 (±0.001)	0.222	0.629 (±0.007)
DeepDTA	0.863 (±0.002)	0.194	0.673 (±0.009)
WideDTA	0.875 (±0.001)	0.179	–
MT-DTI	0.882 (±0.001)	0.152	0.738 (±0.006)
DeepCDA	0.889 (±0.002)	0.176	0.682 (±0.008)
MATT-DTI	0.889 (±0.001)	0.150	0.756 (±0.011)
GraphDTA	0.891 (±0.002)	0.139	–
MultiscaleDTA	**0.893(±0.003)**	**0.135**	**0.793(±0.009)**

3.4. Ablation experiment

To investigate the contribution of each part to the performance of DTA prediction in the proposed MultiscaleDTA, each part from the method MultiscaleDTA is removed. We implement the ablation experiments with the variants of our model by using Davis and KIBA datasets:

- § MultiscaleDTA without the multi-scale information (w/o MSI) only uses the information from the last CNN layer as the representation of the drug or target.
- § MultiscaleDTA without the self-attention mechanism (w/o SAM) directly concatenates the features from different CNN layers as the representation of the drug or target.

Fig. 2 illustrates the comparative results of MultiscaleDTA and its two variants on the two datasets. As shown in Fig. 2, we can see that the proposed MultiscaleDTA achieves better prediction performance than its variants, which indicates that incorporating multi-scale information and the self-attention mechanism can obtain a more discriminative representation. In detail, the performance gap between MultiscaleDTA (w/o MSI) and MultiscaleDTA is the greatest, demonstrating that multi-scale information plays the most important role in our model and removing this part will greatly reduce the model performance. Besides, MultiscaleDTA (w/o MSI) performance is worse than MultiscaleDTA (w/o SAM) since it loses the information from different CNN layers, which contains more sufficient and effective task-related information. The deprecation of the self-attention mechanism also results in the performance reduction. It also indicates the effectiveness of the self-attention mechanism in optimizing the obtained features.

4. Conclusion

Identifying the binding affinity of the unseen drug-target pair is crucial in facilitating drug discovery. In this work, we present MultiscaleDTA, a novel end-to-end deep learning-based method for accurately predicting DTA. Specifically, we first build two multi-scale feature

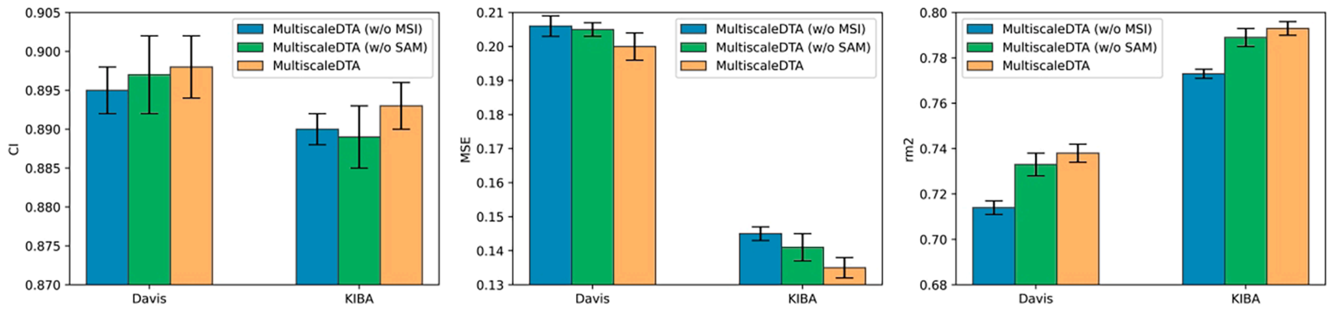

Fig. 2. CI, MSE and r_m^2 results from variants of our method on the Davis and KIBA dataset.

extractors to learn multi-scale features of drug compounds and target proteins, respectively. Next, a self-attention mechanism is introduced to calculate the contribution of each feature from different CNN layers to DTA prediction, which can let our model focus on important parts. Then, the features from different CNN layers are concatenated as the drug (target) representation. Finally, the representations of the drug and target are further concatenated, and the combined features are fed into the fully connected layers to implement binding affinity prediction. Experimental results on two DTA datasets, including KIBA and Davis, show that the proposed MultiscaleDTA is superior to the existing models. In particular, we also use Human and *C.elegans* datasets to evaluate the performance of MultiscaleDTA on the classification task, and the experimental results indicate that our model also achieves competitive performance.

Funding

This study was supported by the Natural Science Foundation of China (No. 62071278).

CRediT authorship contribution statement

Haoyang Chen: Methodology, Writing – original draft, Writing – review & editing. **Dahe Li:** Formal analysis, Data curation. **Jiaqi Liao:** Formal analysis, Investigation. **Lesong Wei:** Conceptualization, Methodology, Validation, Writing - review & editing. **Leyi Wei:** Supervision, Conceptualization, Writing - Review & Editing, Funding acquisition

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] J.W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency, Nat. Rev. Drug Discovery 11 (3) (2012) 191–200.

[2] M.K. Gilson, L. Tiqing, B. Michael, N. George, H. Linda, C. Jenny, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Research (D1) (2016) D1045-D1053.

[3] S. Kim, C. Jie, T. Cheng, A. Gindulyte, E.E. Bolton, PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (Database issue) (2018).

[4] Y. Hu, H. Zhang, B. Liu, S. Gao, T. Wang, Z. Han, P. International Genomics of Alzheimer's, X. Ji, G. Liu, rs34331204 regulates TSPAN13 expression and contributes to Alzheimer's disease with sex differences, Brain 143(11) (2020) e95.

[5] Y. Hu, S. Qiu, L. Cheng, Integration of Multiple-Omics Data to Analyze the Population-Specific Differences for Coronary Artery Disease, Comput. Math Methods Med. 2021 (2021) 7036592.

[6] P.J. Ballester, J.B. Mitchell, A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking, Bioinformatics 26 (9) (2010) 1169–1175.

[7] H. Tong, M. Heidemeyer, F. Ban, A. Cherkasov…,, SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines, J. Cheminform. 9 (1) (2017).

[8] B. Tanoori, M.Z. Jahromi, E.G. Mansoori, Drug-target continuous binding affinity prediction using multiple sources of information, Expert Syst. Appl. 186 (2021), 115810.

[9] F. Boyles, C.M. Deane, G.M. Morris, Learning from the ligand: using ligand-based features to improve binding affinity prediction, Bioinformatics 36 (3) (2019) 758–764.

[10] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J.B. Ghasemi, A. Masoudi-Nejad, DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks, Bioinformatics 36 (17) (2020) 4633–4642.

[11] Y. Zeng, X. Chen, Y. Luo, X. Li, D. Peng, Deep drug-target binding affinity prediction with multiple attention blocks, Briefings Bioinf. 22 (5) (2021) bbab117.

[12] H. Öztürk, E. Ozkirimli, A. Özgür, WideDTA: prediction of drug-target binding affinity, (2019).

[13] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, H.J.J.o.P.R. Lu, Deep-Learning-Based Drug-Target Interaction Prediction, 16(4) (2017) 1401.

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

[15] Y. Hu, J.Y. Sun, Y. Zhang, H. Zhang, S. Gao, T. Wang, Z. Han, L. Wang, B.L. Sun, G. Liu, rs1990622 variant associates with Alzheimer's disease and regulates TMEM106B expression in human brain tissues, BMC Med. 19 (1) (2021) 11.

[16] Y. Hu, Y. Zhang, H. Zhang, S. Gao, L. Wang, T. Wang, Z. Han, P., International Genomics of Alzheimer's, G. Liu, Mendelian randomization highlights causal association between genetically increased C-reactive protein levels and reduced Alzheimer's disease risk, Alzheimers Dement (2022).

[17] Y. Hu, Y. Zhang, H. Zhang, S. Gao, L. Wang, T. Wang, Z. Han, B.L. Sun, G. Liu, Cognitive performance protects against Alzheimer's disease independently of educational attainment and intelligence, Mol. Psychiatry (2022).

[18] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, S.J.B. Venkatesh, GraphDTA: Predicting drug–target binding affinity with graph neural networks, (2020).

[19] Q. Tian, M. Ding, H. Yang, C. Yue, Y. Zhong, Z. Du, D. Liu, J. Liu, Y. Deng, Predicting drug-target affinity based on recurrent neural networks and graph convolutional neural networks, Comb. Chem. High Throughput Screening 25 (4) (2022) 634–641.

[20] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, Z. Wei, Drug–target affinity prediction using graph neural network and contact maps, RSC Adv. 10 (35) (2020) 20701–20712.

[21] M.I. Davis, J.P. Hunt, S. Herrgard, P. Ciceri, L.M. Wodicka, G. Pallares, M. Hocker, D.K. Treiber, P.P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, Nat. Biotechnol. 29 (11) (2011) 1046–1051.

[22] S. Derrick, Making sense of large data sets, EE: evaluation engineering: the magazine of electronic, Evaluation 50 (12) (2011) 18–20.

[23] Z. Hakime, Z. Arzucan, O. Elif, DeepDTA: deep drug-target binding affinity prediction, Bioinformatics 17 (2018) 17.

[24] M. Jiang, Z. Li, S. Zhang, S. Wang, Z. Wei, Drug–target affinity prediction using graph neural network and contact maps, RSC Adv. 10 (35) (2020) 20701–20712.

[25] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, S. Venkatesh, GraphDTA: Predicting drug–target binding affinity with graph neural networks, Bioinformatics (2020).

[26] Z. Yang, W. Zhong, L. Zhao, Y.C. Chen, MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction, Chem. Sci. 13 (2022).

[27] K. Sunghwan, P.A. Thiessen, E.E. Bolton, C. Jie, G. Fu, G. Asta, L. Han, J. He, S. He, B.A. Shoemaker, PubChem Substance and Compound databases, Nucleic Acids Research (D1) (2016) D1202-D1213.

[28] U.P. Consortium, UniProt: a hub for protein information, Nucleic Acids Re. D1 (2015) 204–212.

[29] G. Stefan, K. Michael, D. Mathias, C. Monica, S. Christian, P. Evangelia, A. Jessica, U.E. Garcia, G. Andreas, J.L. Juhl, SuperTarget and Matador: resources for exploring drug-target relationships, Nucl Acids Research 36 (Database issue) (2008).

[30] D.S. Wishart, C. Knox, C.G. An, S. Shrivastava, M. Hassanali, P. Stothard, C. Zhan, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Oxford University Press (90001), 2006.

[31] D. Weininger, Smiles,, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36.

[32] H. Fu, Z. Niu, C. Zhang, J. Ma, J. Chen, Visual Cortex Inspired CNN Model for Feature Construction in Text Analysis, Front. Comput. Neurosci. 10 (2016) 64-.

[33] W. Alam, S.D. Ali, H. Tayara, K.T. Chong, A CNN-Based RNA N6-Methyladenosine Site Predictor for Multiple Species Using Heterogeneous Features Representation, IEEE Access PP(99) (2020) 1-1.

[34] Y. Ma, C. Yan, A Concurrent Neural Network (CNN) Method for RNA-binding Site Prediction. 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019.

[35] K. Liu, G. Kang, 3D multi-view convolutional neural networks for lung nodule classification, PLoS ONE 12 (1) (2017) 12–22.

[36] A.C. Gilbert, Z. Yi, K. Lee, Y. Zhang, H. Lee, Towards Understanding the Invertibility of Convolutional Neural Networks, Twenty-sixth International Joint Conference on Artificial Intelligence, 2017.

[37] G. Levi, T. Hassncer, Age and gender classification using convolutional neural networks, IEEE Conference on Computer Vision & Pattern Recognition Workshops, 2015, pp. 34-42.

[38] G. Mithat, H.J.B. Glenn, Concordance probability and discriminatory power in proportional hazards regression, 92(4) (2005) 965-970.

[39] M. Tsubaki, K. Tomii, J.J.B.-O.-. Sese, Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, (2019).

[40] S. Zheng, Y. Li, S. Chen, J. Xu, Y.J.N.M.I. Yang, Predicting drug–protein interaction using quasi-visual question answering system, 2(2) (2020) 134-140.

[41] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, M.J. B. Zheng, TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments, (2020).