Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.851

ORIGINAL RESEARCH ARTICLE

DFDTA-MultiAtt: Multi-attention based deep learning ensemble fusion network for drug target affinity prediction

Balanand Jha[1,*], Akshay Deepak[1], Vikash Kumar[1], Gopalakrishnan Krishnasamy[2]

[1] *Department of Computer Science and Engineering, National Institute of Technology Patna, Patna 800005, India*

[2] *Department of Mathematics & Computer Science, Central State University, Wilberforce, Ohio 45384, United States*

*** Corresponding author:** Balanand Jha, balanandj.ph21.cs@nitp.ac.in

ABSTRACT

An essential step in the drug development process is the accurate detection of drug-target interactions (DTI). The importance of binding affinity values in understanding protein-ligand interactions was previously disregarded, and DTI prediction was only seen as a binary classification problem. In this regard, we introduced the DFDTA-MultiAtt model for predicting the drug target binding affinity in two stages using the structural and sequential information. The first step of the first stage involves retrieving features from sequence data using a bi-directional long short term memory (Bi-LSTM) architecture together with a multi-attention module and dilated convolutional neural network (dilated-CNN) architecture, and the second step features are learnt from structure representation once again using a dilated-CNN. To predict the binding affinity, the second stage uses an ensemble learning model. The proposed model also produces findings with a greater overall accuracy when compared to contemporary state-of-the-art methods. The model generates an enormous +0.006 concordance index (CI) score on the Davis dataset and reduces the mean square error (MSE) by 0.174 on the KIBA dataset.

Keywords: drug-target interaction; multi-attention; dilated CNN; Bi-LSTM

ARTICLE INFO

Received: 30 June 2023
Accepted: 6 September 2023
Available online: 14 March 2024

1. Introduction

Physiology of the living being is a complex equilibrium maintained by a nexus of chemical activities involving a number of organic and inorganic compounds and biomolecules. Proteins, in their diverse shape and size, are responsible for performing these physiological activities. The activity of proteins, in turn, is affected by the organic and inorganic compounds in their vicinity. These compounds bind to the active or regulatory sites on the proteins thereby changing its activity. Proper physiology of the body is essential for healthy life. However, due to the multiple factors like life style, environmental changes, pathogens, the physiological equilibrium gets disturbed. Drugs are the compounds that are administered to the living beings to bring the imbalanced physiological equilibrium back on track. In order to identify the drug effective for a particular physiological condition, it is essential to identify the protein responsible for the condition as well as its functional mechanism. Drugs usually bind to these proteins and alter their activities in desired ways. These proteins are known as the "targets".

Recently, drug target interaction prediction has been employed as one of the early phases in drug development, drug repurposing, and

drug side effect prediction research. The primary goal here is to identify new compounds that interact with the predetermined targets. Chemical compounds that make up the building blocks of drugs act as a key object in drug discovery, allowing researchers to understand how these compounds interact with the target proteins. Target profiles and therapeutic benefits of a large number of compounds are yet to be discovered. For instance, while there are almost 90 million chemicals in the PubChem library, most of them lack distinct target interaction characteristics.

In addition to assembling several massive raw data sets, researchers investigating chemical compounds have access to enormous amounts of data on a variety of compound features, properties, and target proteins. Due to the time and resource-intensive nature of purely biological experiments, researchers in the field of drug development are turning to in-silico drug-target interaction prediction algorithms. These algorithms enable them to effectively manage and understand the vast, intricate data with high-dimensional complexity and also prove valuable in accelerating the development of new drugs.

Traditionally, the identification of drug targets has been seen as a binary classification problem. Binary classifiers fail to distinguish between high- and low-affinity interactions because they consider all binding events identically. This ignores important information on the variety of affinities that may direct drug development. Beyond this, the gold standard datasets produced by Yamanishi et al. in 2008[1] have been utilized in the great majority of DTI literature performed during the previous 10 years. These datasets comprised combinations of medications and targets with ambiguous binding information; these combinations were referred to as negative samples. In DTI research, however, more recently, datasets with a higher level of realism have started to be used. These datasets are made using predetermined binding threshold values[2]. The predicted values for drug-target binding affinity from the regression model are also useful for finding innovative medicines with the required interaction characteristics.

Identifying a drug's binding affinity to its target is crucial in contemporary biological research. The similarity-based method and feature-based method are two computational techniques that are often used for this purpose. The similarity-based approach is predicated on the idea that equivalent targets are often targeted by similar medicines, which is validated by previous studies[3–5]. As shown in multiple studies[1,6,7], the feature-based technique, on the other hand, entails creating feature vectors using known descriptors of the drug and target.

KronRLS[8] and SimBoost[9] are two well-known statistical machine learning algorithms for predicting drug-target binding affinity. Protein-protein and drug-drug similarity data are fed into the similarity-based KronRLS method[8]. SimBoost[9], in contrast, analyses the drug-target network to assess each drug and target combination and creates drug-target features using similarity networks. It then employs a regression tree model with larger gradients for performing regression tasks.

With the increasing application scenarios in deep learning, deep learning models have shown great success in addressing various bioinformatics problems[10–14] particularly in drug discovery. When it comes to removing characteristics from high-throughput data, deep learning approaches beat statistical machine learning algorithms. DeepDTA[15] builds a three-layer CNN module to predict binding affinity by label embedding drug Simplified Molecular Input Line Entry System (SMILES) and protein sequences. First, it uses two deep CNN-based architectural blocks—one for protein sequences and one for SMILES—to learn the latent properties of the target and drug independently. Finally, the model was tested for binary classification. WideDTA[16] encodes drug SMILES and protein sequences using a text-based technique with four separate textual components. It has a CNN-based architecture and employs four input representations. Deep neural networks are used to estimate the binding affinity using feature information. Maha et al.[17] introduced Affinity2Vec, a regression-based strategy that frames the entire endeavor as a graph-based problem. They use data from several sources, such as drug-drug similarity, target-target similarity, and drug-target binding affinities, to produce a weighted

heterogeneous network. Next, the regression task is carried out using this network. Further, by using the neighbor connection of similarity and sharing to extract features, Ru et al.[18] proposed a ranking framework using regression features that predict affinity values. GANsDTA[19] is a generative adversarial network (GANs)-based technique to predict binding. Utilizing the GANs' generator module to extract features and the regression module for prediction. It gives the model the ability to learn the characteristics of proteins and drugs, improving feature representation. Additionally, state-of-the-art DeepFusionDTA[20] uses an ensemble learning strategy to anticipate regression and analysis module-based deep learning to construct a feature map of a probable protein and drug combination employing sequencing and structural data. This model fails to illustrate the effectiveness of the attention mechanism for this problem as the attention mechanism may help to concentrate on significant molecular interactions for better affinity prediction.

In this study, we use a multi-attention-based deep learning ensemble fusion network called DFDTA-MultiAtt to obtain an approximate binding affinity value. The network is made up of a dilated-CNN block and a Bi-LSTM block with a multi-attention module. With the help of this proposed architecture, it is possible to learn the structural and sequential data needed to predict binding affinity. A neural network is also used to combine feature maps for potential drug-target combinations. Through the employing of multiple modules, the structure and sequence data were examined to build this feature map. Finally, lightGBM is used for predicting the binding affinity utilizing the generated merged feature map.

Additionally, drugs are represented using SMILES sequences and Morgan Fingerprint descriptors whereas target proteins are represented as amino acid sequences and secondary structures, generated using SPIDER3. For learning the sequential information, Bi-LSTM with multi-attention module with dilated-CNN is employed where, as for the structure data, only deep dilated-CNN with varying convolution layer with different dilation rates is employed. This proposed framework provides sequence shorter as well as longer amino acid interaction information while applying more weightage to the relevant amino acid using the multi-attention mechanism in case sequence information. Beyond this, multi-attention module improved drug target affinity prediction by concentrating on significant molecular interactions. It also provides an understanding of the reasons behind predictions and aids better model accuracy in drug development. Moreover, a deep dilated network will produce the local as well as global relations using different dilated rates and deep networks.

The rest of the paper is as follows: In Section 2, proposed method is elaborated. Section 3 consists of experimental details along with dataset descriptions. Section 4 described the produced results and overall comparison with state-of-the-methods. Finally, Section 5 concludes the paper.

2. Material and methods

2.1. Drug-target interaction framework

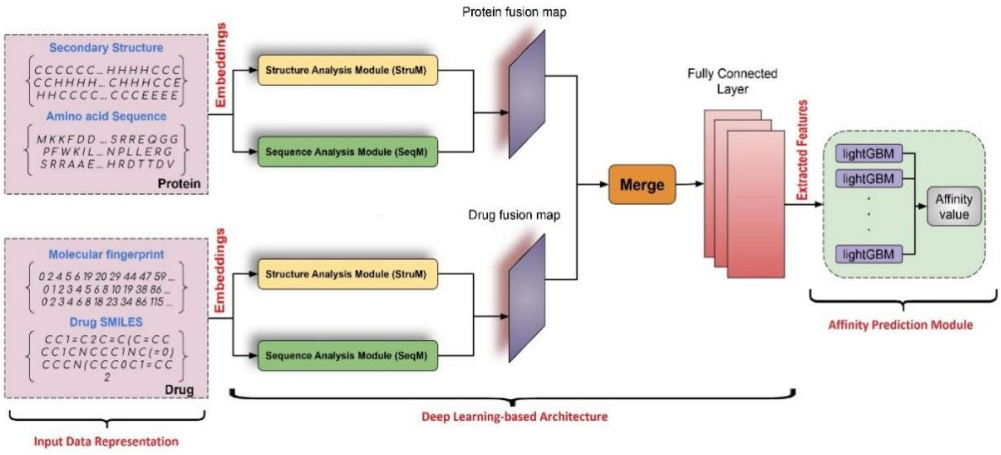

Figure 1. End-to-end framework.

In this work, deep learning-based architectures are employed to generate fusion feature maps for both protein and drug data, encompassing both structure and sequence information. These fused feature maps effectively represent the combinations of drugs and targets. Subsequently, dimensionality reduction techniques are applied to decrease the dimensionality of these fused feature maps. Finally, lightGBM regression module is utilized to determine the binding affinity value utilizing the reduced features. The proposed framework is shown in **Figure 1**, and the different modules comprising the framework are discussed in Sections 2.1 to 2.4.

2.1.1. Input data representation

This is the first phase of the end-to-end framework shown in **Figure 1**. Here, two different pair representations of the dataset are fed for the training of the proposed deep learning-based architecture. These pairs are molecular and protein representation and their effectiveness in combination with drug target interaction prediction are employed. Molecular (drug) representation consists of compound smiles and molecular fingerprint whereas Proteins consists of primary structure—a sequence of amino acids and secondary structure—sequential arrangement of the backbone atoms extracted using SPIDER3. These representations are discussed next in Section 2.2.

2.1.2. Deep learning-based architecture

This is the second phase of this end-to-end framework as shown in **Figure 1**. The proposed architectures take the embedding of data (discussed in Section 2.1.1) as input and produce the final features (represented as Extracted Features in **Figure 1**). For extracting the intermediate features, we have (i) Sequence Analysis Module and (ii) Structure Analysis Module as shown in **Figure 1**. Sequence analysis module consists of two blocks (a) Dilated-CNN block, and (b) Bi-LSTM with multi-attention module as shown in **Figure 2**. Structure analysis module consist of only dilated-CNN block as shown in **Figure 3**. Whereas, sequence analysis module takes sequential information of the protein (sequence) and drug (SMILES) whereas structural analysis module takes structure information of the protein (secondary) and drug (molecular) and produces two sets of feature maps. These feature maps are merged and fed to the dense layer for dimensionality reduction and further fed to the downstream network. These architectures are deeply described in Section 2.3.

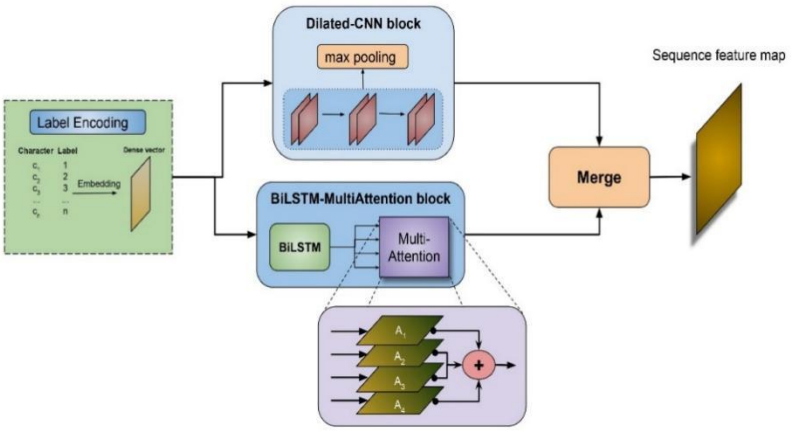

Figure 2. Sequence analysis module.

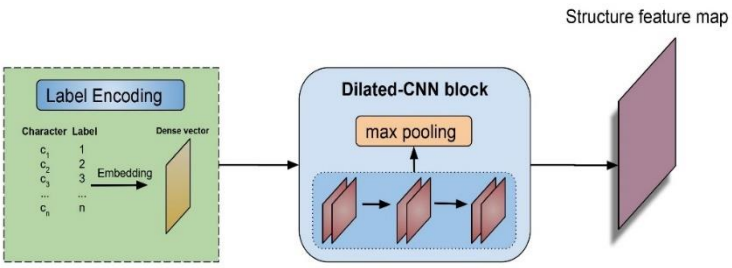

Figure 3. Structure analysis module.

2.1.3. Affinity prediction module

This is the last module mainly used to estimate the binding affinity. Here, an ensemble bagging-based lightGBM regression module receives input as features, which represent the drug-target interaction pair. Further, this module produces the binding affinity value as output. This prediction module is elaborated in Section 2.4.

2.2. Data representation

2.2.1. Molecular representation

Compound SMILES: SMILES (Simplified Molecular Input Line Entry System), represents chemical structures as text, also called compound SMILES. Here, to represent SMILES inputs, we used label encoding, which employs extracted 64 unique letters integers as a category to embed the compound SMILES sequences.

Molecular Fingerprint: A molecule's fingerprint is a numerical representation of its chemical constituents. It provides a more informative feature description and is also easier to analyze the role of sequence and structure in predicting the interaction of drug and target. Here, RDKit[21] is used to generate the Morgan fingerprint. Next, the substructure information is converted into dense feature vector representation using embedding method.

2.2.2. Protein representation

Protein Sequence: Protein Sequence is represented using the chain of essential amino acids. Here, each amino acid is represented by a distinct number, and with the following representation, the label encoding technique is used to embed each protein sequence.

Protein Secondary Structure: The secondary structure is extracted from the protein sequence using SPIDER3, represented as a sequence of characters. The secondary structure of the protein is shown by the three characters "H" for alpha-helix, "E" for beta-strand, and "C" for coil in this depiction. The information about the amino acids' short- and long-range interactions is encoded in each character of the secondary structure sequence. This sequence is then transformed into dense feature representations using label encoding.

2.3. Deep learning-based architecture

This phase takes embeddings (molecular and protein representation) as input and extracts the features using sequence analysis module (shown in **Figure 2**) and structure analysis module (shown in **Figure 3**) and fed to the downstream network after concatenating these features using merge layer.

2.3.1. Dilated-CNN block

Convolutional Neural Networks (CNNs) are effective weight-sharing techniques that perform well in spotting patterns in a variety of contexts. It can acquire local representations with a comparatively small number of parameters. Convolutional layer after stacking allows CNN to collect both local and global data, allowing for a more comprehensive interpretation of the data. It's vital to keep in mind, too, that adding more convolutional layers also increases the overall number of trainable parameters. So, we have incorporated the dilated-CNN block which learns short- and long-range interaction with varying dilation rates and lowers the overall trainable parameter[22].

In order to capture both local and global interactions while minimizing the number of parameters, a deep dilated-CNN architecture made up of three layers of dilated convolution is used in this work. The dilated convolution operates by translating the input region to the filter with dilation constant, k. Since each input unit is separated from neighboring units in both vertical-and-horizontal dimensions, there is a wider receptive field and the capacity to integrate more contextual information. The receptive field are increased using different dilation rate, and convolution operation is calculated as Equation (1).

$$ConV(X)_{i,k} = LeakyReLU\left\{\sum_{m=0}^{M-1} W_m \cdot X_{i.step+k}\right\} \quad (1)$$

where M is the length of the filter, and k is the dilation rate.

Next, the outputs are given to non-linear activation function (Equation (2)) for learning the complex information.

$$LeakyReLU(x) = \begin{cases} x, & x \geq 0 \\ \frac{x}{a}, & x < 0 \end{cases} \quad (2)$$

As output of this block, it produces important features context across larger receptive field and also minimize the overall trainable parameter.

2.3.2. Bi-LSTM block

Bidirectional LSTM is another sequence processing model, which learns the interaction in both forward and backward directions. Forward LSTM learns the interaction in forward direction whereas Backward LSTM learns the interaction from backward direction. Additionally, it gains knowledge of short- and long-range interactions, efficiently enhances the network's data availability, and broadens the context. Bi-LSTM is more suited to sequence information on drugs and targets since it can capture long-term dependence and identify contextual links. There are three gates in each LSTM unit: (i) an input gate, (ii) a forget gate, and, (iii) an output gate. These gates are stated as follows:

$$f_t = \sigma(W_f x_t + U_f b_{\{t-1\}} + b_f)$$
$$i_t = \sigma(W_i x_t + U_i b_{\{t-1\}} + b_i)$$
$$o_t = \sigma(W_o x_t + U_o b_{\{t-1\}} + b_o)$$
$$C_t = i_t \cdot tanh\left(W_c x_t + U_c b_{\{t-1\}} + b_c\right) + f_t \cdot C_{\{t-1\}}$$
$$h_t = o_t \cdot tanh(C_t)$$

where W, U are parameter matrices and b is bias.

Here, input gate 'i_t' add more relevant information to the current memory state 'C_t', whereas forget gate 'f_t' removes the irrelevant information from the memory state and the output gate 'o_t' produces the selective output to increase the efficiency of the model.

2.3.3. Multi-attention module

The attention mechanism is a technique that can assist a neural network in memorizing extended sequences of information or data. Here we are using multi-attention module. A multi attention module is a group of M-attention layers. In our experiment, we have considered (M = 4) four attention layers and summed up the output of each attention layer. It takes input from the Bi-LSTM and assign the weights to input. The output of multi-attention is concatenated with the output of Bi-LSTM for better feature mapping as all the inputs are responsible for mapping to corresponding output. The architecture of the multi-attention module is shown in **Figure 1**.

The attention layer calculates set of attention weights denoted by $\alpha(t_1), \alpha(t_2), \ldots, \alpha(t)$. The context vector c_i for the output word y_i is generated using the weighted sum as shown in Equation (3).

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij}\, h_j \quad (3)$$

By normalizing the output score defined by the function that captures the alignment between input at j and output at i the attention weights are generated as shown in Equation (4).

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_z} \exp\left(e_{ik}\right)} \tag{4}$$

$$e_{ij} = a_{ij}(s_{i-1}, h_j) \tag{5}$$

Finally, the attention weights are multiplied with the output of the Bi-LSTM and the generated output (shown in Equation (5)) is given downstream network i.e., merge layer.

2.3.4. Merge module

In **Figure 1**, Sequence analyses module consists of two deep learning-based architectures, (i) Bi-LSTM with multi-attention layer and (ii) Dilated-CNN block. Both short- and long-range interaction information is captured using Bi-LSTM and dilated-CNN block. Bi-LSTM captures temporal information whereas dilated-CNN extracts local patterns and correlations from sequential data. Both the architecture is trained parallels, so merge the output of both the architecture, a merge layer is used. Merge layer merges the output of both the architecture in sequence analyses module and final feature map is given as output from this module.

Again, the output of each analysis module is combined using a merge layer to create the final feature maps, which include data on the protein and drug combination. Therefore, high dimensional features are generated in the final feature map after the final merging layer. Therefore, three dense connected layers with 1024, 1024, and 512 neurons on each layer are used to map the features to the lower dimensions. The regression module uses the output of the last layer to estimate affinities.

2.4. Affinity prediction module

Regression performance can be increased by integrating different weak learning methods, which has two key advantages. Due to the limited quantity of data, the real distribution of data is often impossible to describe with a single hypothesis. Therefore, weighing many hypothetical models can aid in lowering the probability of making the wrong choice. Instead of using a single model-based approach, such as a decision tree or neural network architecture, the ensemble approach achieves regression value by combining a number of weak models from several starting points. Here, we use ensemble learning to increase the regression's accuracy and predict binding affinity values using the fully connected layer's 512-dimensional feature vector. Bagging approach averages the results for the final prediction after combining the outputs from numerous lightGBM models produced by diverse subsets of the input characteristics. LightGBM is excellent in fusing weak and simple learners into a strong learner.

3. Experiment Details

In this section, we discuss the benchmarked datasets used to evaluate the proposed model. The benchmarked datasets are described in Section 3.1, and the evaluation metric are discussed in Section 3.2. We have implemented the proposed model using Python with Keras and TensorFlow. The efficiency of the various strategy described in this study are examined. Finally, we compared our results with the contemporary state-of-the-art methods.

3.1. Dataset

We have used two benchmarked binding affinity prediction datasets, (i) Davis[22], and (ii) KIBA[8]. The detailed description of these two datasets is given in the **Table 1**. Davis[22] consists of 68 drugs and 442 proteins whereas KIBA[8] consists of 2111 drugs and 229 proteins as given in **Table 1**. **Figure 4** shows the frequency distribution between the drugs and their length and protein sequences and their length. So, Davis consists of a maximum drug length of 103 and 64 as the average length whereas protein sequences consist of maximum 2549 amino acids and 788 amino acids as average length. Similarly, the KIBA dataset has maximum drug length of 590 and 58 as average length whereas protein sequences consist of maximum 4128 amino acids and

728 amino acids as average length.

Moreover, the Davis dataset includes selectivity assays for the kinase protein-family as well as the relevant inhibitors, along-with their dissociation constant K_d values. The KIBA dataset, on the other hand, came from a method known as KIBA, which included kinase inhibitor bioactivities from several data sources such as K_i, K_d, and IC50[8]. By utilizing the statistical information contained in K_i, K_d, and IC50, KIBA scores were created to optimize the degree of coherence amongst them.

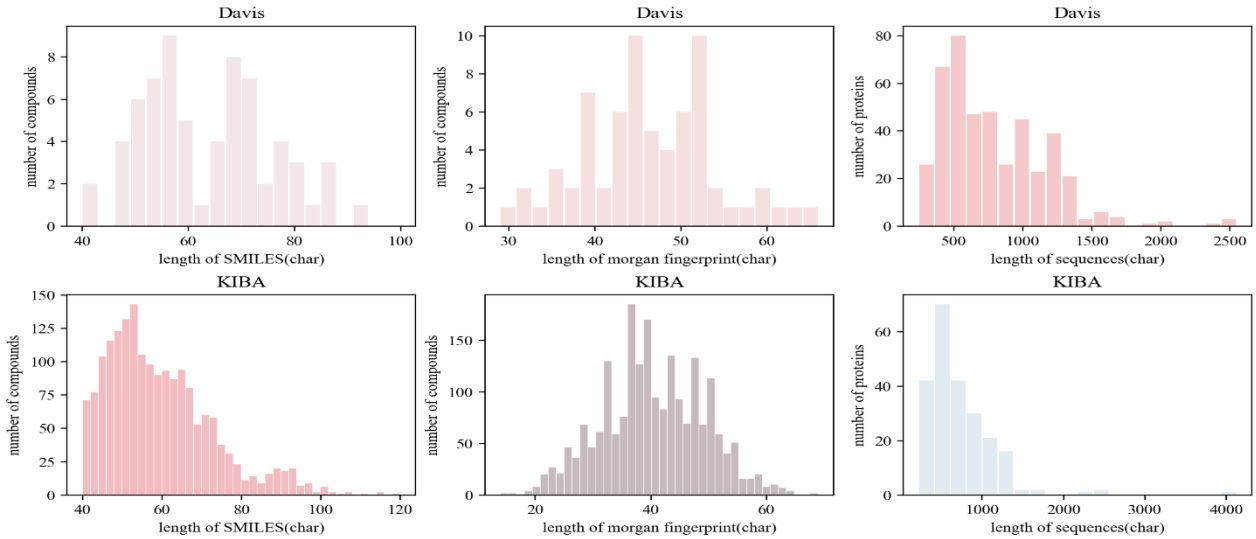

Figure 4. Analysis of the datasets.

Table 1. Dataset description.

Dataset	Protein	Compound	Interactions
Davis (K_d)	442	68	30,056
KIBA	229	2111	118,254

In contrast to Pahikkala et al.[23], where the binding affinity score of the Davis dataset were used directly as K_d values, our approach involves transforming these values into log space, known as pK_d similar to the methodology employed in He et al.[9]. This transformation is outlined in Equation (6).

$$pK_d = -\log_{10}\left(\frac{K_d}{1e9}\right) \tag{6}$$

As illustrated in **Figure 4**, the protein sequence may be as long as 4128 characters, with an average length of 728 characters, whereas the longest drug SMILES in the KIBA dataset had a maximum length of 590 characters and an average length of 58 characters.

3.2. Evaluation matrices

Affinity prediction comes in category of regression problem. In this regard, we used two well-known evaluation metrics to evaluate our proposed architecture: (i) mean squared error and (ii) concordance index. These are briefly described below.

3.2.1. Mean squared error

In regression tasks, MSE is widely used to determine how accurately the predicted values are differentiated from the actual values. Equation (7) illustrates the MSE formula, where P stands for the anticipated output, Y for the actual outputs, and n for the sample size. Performance of the regression modal increases as MSE drops.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - Y_i)^2 \quad (7)$$

3.2.2. Concordance index

A variant of the receiver operating characteristic curve is the concordance index, often known as the C-index or CI. It is a summary evaluation of the model's ability to distinguish between various drug-target combinations and score them appropriately. Using a set of data's concordance index, which takes into account each data pair independently, one may estimate the probability that the predicted labels are in the same order as the real label values. Equation (8) displays the CI formula as follows:

$$C_i = \frac{1}{Z}\sum_{\delta_x > \delta_y} s(b_x - b_y) \quad (8)$$

where b_x is the predicted value with respect to higher affinity δx, b_y is the predicted value against the higher affinity δy, Z is standardization constant, $s(b)$ is the step function shown in Equation 9.

$$s(b) = \begin{cases} 1, & if\ b > 0 \\ 0.5, & if\ b = 0 \\ 0, & if\ b < 0 \end{cases} \quad (9)$$

3.3. Baselines

We have compared our work with the contemporary state-of-the-art approaches, namely: KronRLS[8], SimBoost[9], DeepDTA[15], WideDTA[16] and DeepFusionDTA[20].

3.3.1. KronRLS

The regularized least squares model (RLS)-based method KronRLS[8] can predict binding values that may be binary or continuous. It trains the function $\boldsymbol{f(x)}$ for all the probable drug-target pairs $\boldsymbol{x} \in \{\boldsymbol{d_i} \times \boldsymbol{t_j}\}$. Here, $\boldsymbol{d_i}$ and $\boldsymbol{t_j}$ denote the set of chemical compounds and their target biological receptors respectively. For calculating the difference between two pairs, one must learn an objective function. Therefore, $\boldsymbol{f}$ is framed as the problem of optimizing the minimizer of an associated objective function.

3.3.2. SimBoost

For the purpose of creating features for each drug, target, and drug-target pair, SimBoost[9] is a gradient enhancement machine-based technique. Each drug-target pair has a feature vector associated with it in SimBoost. Drug-drug resemblance, target-target analogy, and drug-target binding are the three networks that SimBoost employs. In the latter case, a node may represent a drug as well as a target, and binding affinity values connect the drug and target nodes. Latent vectors obtained through matrix-factorization are also used in this network. Gradient boosting regression trees, a supervised learning technique, are utilized for DTI prediction.

3.3.3. DeepDTA

A deep learning-based model that uses CNN-based architecture is called DeepDTA[15]. To learn the latent features of the target and drug separately, it employs two deep CNN-based architecture blocks—one for protein sequences and one for SMILES—before pooling layers are used to reduce the size. The KIBA and Davis datasets were used to assess DeepDTA's performance. Despite being designed for continuous value prediction; it was evaluated for binary classification.

3.3.4. WideDTA

In order to improve DeepDTA results, WideDTA[16] additionally makes use of CNN-based architecture and four input representations. Drug SMILES and protein sequences are represented by WideDTA as word groupings rather than whole sequences. Drugs and proteins often have word counts of three and eight, respectively. From each of the input data, features are extracted using four identical models. To predict binding

affinity, the collected characteristics from each model are aggregated and supplied into downstream densely connected network.

3.3.5. DeepFusionDTA

To carry out Drug-target identification, DeepFusionDTA[20] employs a two-step design. An analysis module-based deep learning fusion feature map of a potential protein and drug combination is initially created using sequencing and structural data. The second step involves using a bagging-based ensemble learning approach to forecast regression. For assessment, it additionally makes use of the Davis and KIBA datasets.

3.4. Hyperparameter dtails

The feature extraction phase uses grid-search to find the ideal filter size parameters for proteins and medicines because their input lengths differ. We choose two lengths for chemicals and proteins in the grid search, i.e., 4 and 8 for chemicals and 8 and 12 for proteins. Three 1D dilated-CNN with corresponding number of filters as 32, 64, and 96 respectively with varying dilation rate as 2, 3, and 4 are used. Utilizing an averaging method, 48 hidden units are employed to preserve the Bi-LSTM/Bi-GRU layer and the final dimension of the convolution layer while concatenating the output of the convolution layer.

Table 2. The average CI and MSE scores using the different hyper parameter of the test Davis dataset. (RNN – Recurrent Neural Network)

S.No.	RNN	D-CNN	#Attention Layer	CI	MSE
1	Bi-LSTM	1,1,1	1-Att layer	0.884	0.255
2	Bi-GRU	1,1,1	1-Att layer	0.888	0.248
3	Bi-LSTM	2,3,4	1-Att layer	0.887	0.247
4	Bi-LSTM	2,3,4	4-Att layer	0.893	0.247

Four attention layers are utilized to account for the inputs' varying degrees of relevance. The bagging-based lightGBM module section builds 800 base learners, with a learning rate of 0.01, and trains the model using 80 leaves. The maximum depth is 150 in order to prevent overfitting. We use 800 estimators to do regression while bagging learning.

Table 3. The average CI and MSE scores using the different hyper parameter of the test KIBA dataset. (RNN – Recurrent Neural Network)

S.No.	RNN	D-CNN	#Attention Layer	CI	MSE
1	Bi-LSTM	1,1,1	1-Att layer	0.854	0.002
2	Bi-GRU	1,1,1	1-Att layer	0.865	0.002
3	Bi-LSTM	2,3,4	1-Att layer	0.864	0.002
4	Bi-LSTM	2,3,4	4-Att layer	0.874	0.002

4. Results and discussion

On the Davis and KIBA datasets, respectively, Tables 2 and 3 present the findings we obtained using the suggested technique with various parameter settings. For the Davis dataset findings provided in Table 2, the top-performing models are the Bi-LSTM with 4-attention layer and the dilated-CNN with dilatation rates of 2, 3, and 4. Our model yields the result of 0.893 CI and 0.247 MSE. On the KIBA dataset, results are reported in Table 3 and all models have the same MSE value, which is very low and the best CI value found is 0.874. By substituting the Bi-LSTM with Bi-GRU and applying a different dilation rate to the various dilated-CNN architectures, as shown in Tables 2 and 3, we have additionally assessed the suggested model. Then, we contrasted the outcomes with cutting-edge techniques.

The result of the proposed model, shown (as asterisk (*)) in Table 4 and Table 5, named as DFDTA-Att*

and DFDTA-MultiAtt*, are compared with the contemporary state-of-the-art methods. Here DFDTA-Att* is based on single attention whereas DFDTA-MultiAtt* is based on multi-attention with 4 attention layers. The proposed model attains lowest MSE score on both the dataset and highest CI score on Davis dataset whereas comparable CI score on KIBA dataset. The MSE is dropped to a significant value due to the addition of multi-attention module in the end-to-end architecture as it concentrates on significant molecular interactions.

Table 4. The average CI and MSE scores of the test Davis dataset (The proposed model is indicated with asterisk (*) and best values are in bold style, Rep – Representation, PS&PDM – Protein sequence & Protein domain motif, and LS&LMCS – Ligand smile & Ligand Maximum Common Substructure).

Models	Target Rep	Drug Rep	CI	MSE
KronRLS[8]	Smith-Waterman	Pubchem-Sim	0.871	0.379
SimBoost[9]	Smith-Waterman	Pubchem-Sim	0.872	0.282
DeepDTA[15]	Smith-Waterman	Pubchem-Sim	0.790	0.608
DeepDTA[15]	Smith-Waterman	CNN	0.886	0.420
DeepDTA[15]	CNN	Pubchem-Sim	0.835	0.419
DeepDTA[15]	CNN	CNN	0.878	0.261
WideDTA[16]	PS&PDM	LS&LMCS	0.886	0.262
DeepFusionDTA[20]	SeqM&StruM	SeqM&StruM	0.887	0.253
DFDTA-Att*	SeqM&StruM	SeqM&StruM	0.885	0.255
DFDTA-MultiAtt*	SeqM&StruM	SeqM&StruM	0.893	0.247

Table 5. The average CI and MSE scores of the test KIBA (The proposed model is indicated with asterisk (*) and best values are in bold style, Rep – Representation, PS&PDM – Protein sequence & Protein domain motif, and LS&LMCS – Ligand smile & Ligand Maximum Common Substructure).

Models	Target Rep	Drug Rep	CI	MSE
KronRLS[8]	Smith-Waterman	Pubchem-Sim	0.782	0.411
SimBoost[9]	Smith-Waterman	Pubchem-Sim	0.836	0.222
DeepDTA[15]	Smith-Waterman	Pubchem-Sim	0.710	0.502
DeepDTA[15]	Smith-Waterman	CNN	0.854	0.204
DeepDTA[15]	CNN	Pubchem-Sim	0.718	0.571
DeepDTA[15]	CNN	CNN	0.863	0.194
WideDTA[16]	PS&PDM	LS&LMCS	0.875	0.179
DeepFusionDTA[20]	SeqM&StruM	SeqM&StruM	0.876	0.176
DFDTA-Att*	SeqM&StruM	SeqM&StruM	0.866	0.002
DFDTA-MultiAtt*	SeqM&StruM	SeqM&StruM	0.874	0.002

5. Conclusion

We proposed a deep learning-based architecture to predict the drug and target binding affinity using structural, protein and drug sequence, and similar data. We use two concurrent deep learning-based architectures for the sequence feature. It uses a dilated-CNN block and Bi-LSTM with a multi-attention module. Also used for structural data is a single dilated-CNN block. For the purpose of forecasting the binding affinity, the acquired characteristics are combined and given to the lightGBM regression model after dimension reduction. In order to compare the performance of the recommended model, we additionally make use of various parameters during feature selection in end-to-end framework.

Our findings show that multi-attention aids in producing better feature maps, which improves accuracy. Additionally, when the dilation rate rises, the receptive field expands and interaction data from farther distances helps to provide a better Concordance index and reduce mean square error.

Author contributions

Conceptualization, BJ and AD; methodology, BJ and AD; software, VK; validation, BJ, AD and GK; formal analysis, BJ; investigation, BJ and VK; resources, BJ and GK; data curation, BJ and VK; writing, BJ and VK; original draft preparation, BJ and AD; writing—review and editing, BJ, AD and GK; visualization, BJ, and VK; supervision, AD and GK; project administration, AD and GK; funding acquisition, AD. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Zhao L, Wang J, Pang L, et al. GANsDTA: Predicting drug-target binding affinity using GANs. Frontiers in Genetics 2020; 10: 1243. doi: 10.3389/fgene.2019.01243
2. Wan F, Zeng J. Deep learning with feature embedding for compound-protein interaction prediction. bioRxiv 2016; 033–086. doi: 10.1101/086033
3. Atias N, Sharan R. An algorithmic framework for predicting side effects of drugs. Journal of Computational Biology 2011; 18(3): 207–218. doi: 10.1089/cmb.2010.0255
4. Öztürk H, Ozkirimli E, Özgür A. A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. BMC Bioinformatics 2016; 17(1): 128. doi: 10.1186/s12859-016-0977-x
5. Iorio F, Tagliaferri R, di Bernardo D. Identifying network of drug mode of action by gene expression profiling. Journal of Computational Biology 2009; 16(2): 241–251. doi: 10.1089/cmb.2008.10TT
6. Keum J, Nam H. Self-BLM: Prediction of drug-target interactions via self-training SVM. PloS One 2017; 12(2): e0171839. doi: 10.1371/journal.pone.0171839
7. Cao DS, Zhang LX, Tan GS, et al. Computational prediction of drug—Target interactions using chemical, biological, and network features. Molecular Informatics 2014; 33(10): 669–681. doi: 10.1002/minf.201400009
8. Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug-target interaction predictions. Brief Bioinformatics 2015; 16(2): 325–337. doi: 10.1093/bib/bbu010
9. He T, Heidemeyer M, Ban F, et al. SimBoost: A read-across approach for predicting drug-target binding affinities using gradient boosting machines. Journal of Cheminformatics 2017; 9(1): 24. doi: 10.1186/s13321-017-0209-z
10. Li J, Pu Y, Tang J, et al. DeepAVP: A dual-channel deep neural network for identifying variable-length antiviral peptides. IEEE Journal of Biomedical and Health Informatics 2020; 24(10): 3012–3019. doi: 10.1109/JBHI.2020.2977091
11. Su R, Liu X, Xiao G, et al. Meta-GDBP: A high-level stacked regression model to improve anticancer drug response prediction. Briefings in Bioinformatics 2020; 21(3): 996–1005. doi: 10.1093/bib/bbz022
12. Su R, Wu H, Xu B, et al. Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2019; 16(4): 1231–1239. doi: 10.1109/TCBB.2018.2858756
13. Li J, Pu Y, Tang J, et al. DeepATT: A hybrid category attention neural network for identifying functional effects of DNA sequences. Briefings in Bioinformatics 2021; 22(3): bbaa159. doi: 10.1093/bib/bbaa159
14. Wei L, Zhou C, Chen H, et al. ACPred-FL: A sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. Bioinformatics 2018; 34(23): 4007–4016. doi: 10.1093/bioinformatics/bty451
15. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: Deep drug-target binding affinity prediction. Bioinformatics 2018; 34(17): 821–829. doi: 10.1093/bioinformatics/bty593
16. Öztürk H, Ozkirimli E, Özgür A. WideDTA: Prediction of drug-target binding affinity. arXiv 2019; arXiv:1902.04166v1. doi: 10.48550/arXiv.1902.04166
17. Ru X, Ye X, Sakurai T, Zou Q. NerLTR-DTA: Drug-target binding affinity prediction based on neighbor relationship and learning to rank. Bioinformatics 2022; 38(7): 1964–1971. doi: 10.1093/bioinformatics/btac048
18. Xiaoqing R, Quan Z, Chen L, Optimization of drug–target affinity prediction methods through feature processing schemes, Bioinformatics, Volume 39, Issue 11, November 2023, btad615
19. Kumar V, Deepak A, Ranjan A, Prakash A. Lite-SeqCNN: A light-weight deep CNN architecture for protein function prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2023; 20(3): 2242–2253. doi: 10.1109/TCBB.2023.3240169
20. 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 Transactions on Computational Biology and Bioinformatics 2022; 19(5): 2760–2769. doi: 10.1109/TCBB.2021.3103966
21. Landrum G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling.

Available online: https://docplayer.net/11897218-Rdkit-a-software-suite-for-cheminformatics-computational-chemistry-and-predictive-modeling.html (accessed on 19 August 2023).
22. Thafar MA, Alshahrani M, Albaradei S, et al. Affinity2Vec: Drug-target binding affinity prediction through representation learning, graph mining, and machine learning. Scientific Reports 2022; 12(1): 4751. doi: 10.1038/s41598-022-08787-9
23. Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug-target interaction predictions. Briefings in Bioinformatics 2015; 16(2): 325–337. doi: 10.1093/bib/bbu010
24. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. Nature Biotechnology 2011; 29(11): 1046–1051. doi: 10.1038/nbt.1990
25. Rogers D, Hahn M. Extended-connectivity fingerprints. Journal of Chemical Information and Modeling 2010; 50(5): 742–754. doi: 10.1021/ci100050t
26. Tang J, Szwajda A, Shakyawar S, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. Journal of Chemical Information and Modeling 2014; 54(3): 735–743. doi: 10.1021/ci400709d