



Article

CSatDTA: Prediction of Drug–Target Binding Affinity Using Convolution Model with Self-Attention

Ashutosh Ghimire ^{1,†} , Hilal Tayara ^{2,†} , Zhenyu Xuan ^{3,*} and Kil To Chong ^{1,4,*}

¹ Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Korea; ashutosh@jbnu.ac.kr

² School of International Engineering and Science, Jeonbuk National University, Jeonju 54896, Korea; hilaltayara@jbnu.ac.kr

³ Department of Biological Sciences, The University of Texas at Dallas, Richardson, TX 75080, USA

⁴ Advanced Electronics and Information Research Center, Jeonbuk National University, Jeonju 54896, Korea

* Correspondence: zhenyu.xuan@utdallas.edu (Z.X.); kitchong@jbnu.ac.kr (K.T.C.)

† These authors contributed equally to this work.

Abstract: Drug discovery, which aids to identify potential novel treatments, entails a broad range of fields of science, including chemistry, pharmacology, and biology. In the early stages of drug development, predicting drug–target affinity is crucial. The proposed model, the prediction of drug–target affinity using a convolution model with self-attention (CSatDTA), applies convolution-based self-attention mechanisms to the molecular drug and target sequences to predict drug–target affinity (DTA) effectively, unlike previous convolution methods, which exhibit significant limitations related to this aspect. The convolutional neural network (CNN) only works on a particular region of information, excluding comprehensive details. Self-attention, on the other hand, is a relatively recent technique for capturing long-range interactions that has been used primarily in sequence modeling tasks. The results of comparative experiments show that CSatDTA surpasses previous sequence-based or other approaches and has outstanding retention abilities.

Keywords: drug–target interaction; binding affinity; attention; convolution neural network; deep learning; artificial intelligence; pharmacometrics; drug discovery and development; proteins; ligands



Citation: Ghimire, A.; Tayara, H.; Xuan, A.; Chong K.T. CSatDTA: Prediction of Drug–Target Binding Affinity Using Convolution Model with Self-Attention. *Int. J. Mol. Sci.* **2022**, *23*, 8453. <https://doi.org/10.3390/ijms23158453>

Academic Editor: George Mihai Nitulescu

Received: 14 July 2022

Accepted: 27 July 2022

Published: 30 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Biochemically dysfunctional allosteric changes in proteins are frequently the cause of many diseases. A drug can change the way native proteins operate in the body, resulting in a desirable therapeutic effect [1]. However, the response of individual patients to a drug varies depending on genetic factors, and prescribers must recognize the need to monitor the outcomes of their prescription [2]. In addition, developing a new drug is expensive [3], and obtaining FDA clearance might require several years [4]. Because of the high cost of drug development, several pharmaceutical companies employ consumer research techniques similar to those used in other industries, in which the genuine value of a product is determined by its consumers [5]. The reduction in drug development costs will not only result in lower drug prices and healthcare expenses for patients, but also allow corporations to produce tailored drugs based on the genetics of individuals. Analyzing the vast extent of available drug and target data in existing databases, emerging and revolutionary computer technologies and deep learning concepts can lower drug development expenses. Currently, neural networks are considered to be relatively beneficial in bioinformatics applications [6–10].

The drug–target interaction identification is an essential aspect of the development of genetic drugs [11], as only the drugs and targets with similar molecular configurations are compatible [12]. The strength of the drug–target interaction pair was measured using DTA. DTA prediction is the essential phase in the computer-aided design (CAD) of a drug, which can accelerate drug development and limit the usage of resources. Drugs can affect

medical conditions by attaching to proteins and can be beneficial or badly affect their functionality. The high binding affinity of a small molecule to a target protein is one factor in selecting a candidate component for drug development. Since the invention of deep learning, its application to DTA prediction to enhance accuracy has received considerable research attention [13]. The equilibrium dissociation constant (K_d) was utilized to evaluate and rank the strength of the interaction binding affinity. Besides K_d , binding affinity is expressed using the half maximal inhibitory concentration (IC_{50}) or the inhibition constant (K_i). The low (K_i) values are associated with the high binding affinity of the ligand to its target. Similarly, the weak attraction and binding between the target molecule and ligand result in a high K_d value [14]. In addition, (IC_{50}) values indicate high binding, which is dependent on the ligand and target concentrations [15].

Several studies have predicted DTA using various models for compounds and targets [16–18]. Deep learning models, most notably CNNs, which learn from protein–ligand composites' three-dimensional (3D) organization, have also been used to score protein–ligand interactions. Some studies [19] have used the Smith–Waterman algorithm, which only employs similarity representations of the targets, and the KronRLS method two-dimensional (2D) compound similarity interpretation of the drugs. Because these techniques require the computation of the similarity matrix, they are constrained to known drug–target composite structures based on the 25,000 drugs in the lists of the Protein Data Bank (PDB) [20], thereby limiting the number of molecules used in the training process. Another study [21] developed a model, referred to as DeepDTA, using a one-dimensional (1D) representation of proteins and ligands to solve these constraints. Instead of using binding complexes' 3D structures or the exterior characteristics, they applied protein sequences and compound representations using the Simplified Molecular Input Line Entry System (SMILES). Additionally, CNN blocks were used to detect bases from SMILES strings and raw protein sequences.

The convolutional neural network [22–25] is a widely used deep learning model that has already achieved significant progress in the realm of feature extraction. Despite possessing the potential benefits of a CNN-based technique, DeepDTA has severe limitations. Long-distance dependencies, for example, are difficult to capture with CNNs. Furthermore, given a raw molecular sequence, CNNs are unable to represent the possible interactions between distant atoms. The CNN model, for example, can record connections between atoms in a series of approximately 35 distances using three convolution layers, each with a filter size of 12. After extracting features from the convolution layers, we used the self-attention mechanism to capture the relationships between atoms in a sequence in our suggested model. Because a multi-head technique with each head corresponding to a feature subspace was used, the applied self-attention mechanism may attend to characteristic subspaces and spatial subspaces simultaneously, in contrast to previous approaches. The proposed model was tested on the popular standard datasets, Davis [26] and KiBA [27].

In contrast to prior deep learning models, our proposed technique results in the elucidation of a more accurate molecular connection. Furthermore, the attention mechanism was used to understand a molecule's high-dimensional structure from a raw sequence. Our suggested technique surpasses existing approaches in terms of DTA prediction without depending on the complex's 3D structure or a 2D representation of the molecule, according to the results. Our findings imply that the attention mechanism is more accurate in terms of abstract and meaningful concepts. Furthermore, we discuss the effective utilization of these discoveries for the further advancement of related research.

2. Result

We consider DeepDTA as the baseline method because it is a typical computational-structure-based approach devised to predict DTAs using a CNN model. To compare our model to up-to-the-minute DeepDTA [21], KronRLS [28], SimBoost [29], WideDTA [30], and GraphDTA [31] models, we adopted the same KiBA [27] and Davis [26] dataset benchmarks. The Davis dataset's binding intensities for 442 targets and 72 drugs were measured using K_d constants varying between 5.00 and 10.80. KiBA is a database that

holds the binding interactions of 2116 drugs and 229 targets, which are represented by KiBA scores that vary between 0.00 and 17.20. To make the evaluation as unbiased as possible, we employed the same number of testing and training cases, as well as the same evaluation measures: CI, MSE, and RMSE. Small MSE and large CI values are ideal. The performance measures originally published in DeepDTA for all the baseline approaches are provided. Table 1 summarizes the hyperparameters used in the experiments. Multiple times, the hyperparameters were fine-tuned.

Table 1. Hyperparameters for CSatDTA.

Hyperparameters	
Learning rate (initially)	0.001
Batch size	64
Optimizer	Adadelta
Kernel initializer	Glorot Normal
CNN layers	2
Attention layers	2
Number of attention heads for SMILES	4
Number of attention heads for proteins	10
Filters for keys and values for SMILES	2
Filters for keys and values for proteins	5

Our CSatDTA model predicts drug–target binding capacity using just drug and protein sequence information. The average MSE, RMSE, and CI values for the datasets KiBA and Davis are shown in Tables 2 and 3, respectively. Previously, the best CI for the Davis dataset was obtained from the GAT-GCN model [31], which was 0.881. Similarly, the same model obtained a better MSE and RMSE, which were 0.245 and 0.494. The CI obtained from our model was 0.892, which is better than the previous performance on the Davis benchmark dataset. The MSE and RMSE were also better than the prior models, with values of 0.241 and 0.490, respectively. The best CI, MSE, and RMSE among the preceding models for the KiBA dataset were 0.891, 0.140, and 0.374, respectively, attained by the GAT-GCN model [31].

Table 2. Performance on the KiBA dataset in terms of prediction.

Method	Compound Rep.	Protein Rep.	MSE	RMSE	CI
KronRLS	Pubchem-Sim	Smith–Waterman	0.411	0.641	0.782
SimBoost	Pubchem-Sim	Smith–Waterman	0.222	0.471	0.836
DeepDTA	1D	1D	0.179	0.423	0.863
WideDTA	1D + LMCS	1D + PDM	0.194	0.440	0.875
GAT_GCN	Graph	1D	0.140	0.374	0.891
CsatDTA (Proposed)	1D	1D	0.134	0.366	0.898

Table 3. Performance on the Davis dataset in terms of prediction.

Method	Compound Rep.	Protein Rep.	MSE	RMSE	CI
KronRLS	Pubchem-Sim	Smith–Waterman	0.379	0.615	0.871
SimBoost	Pubchem-Sim	Smith–Waterman	0.282	0.531	0.872
DeepDTA	1D	1D	0.261	0.510	0.878
WideDTA	1D + LMCS	1D + PDM	0.262	0.511	0.886
GAT_GCN	Graph	1D	0.245	0.494	0.881
CsatDTA (Proposed)	1D	1D	0.241	0.490	0.892

Our model, however, outperformed the data for the KiBA dataset, with a CI, MSE, and RMSE of 0.898, 0.134, and 0.366, respectively. Because of its more comprehensive pharmacological library, the KiBA dataset outperformed the Davis datasets in this analysis.

Moreover, this comes from making the strides as the dataset grows in size, since the KiBA dataset is four-times the size of the Davis dataset. The capacity of deep learning models to capture hidden knowledge improves as the amount of data increases. Figures 1 and 2 depict scatter plots for the Davis and KiBA datasets, respectively.

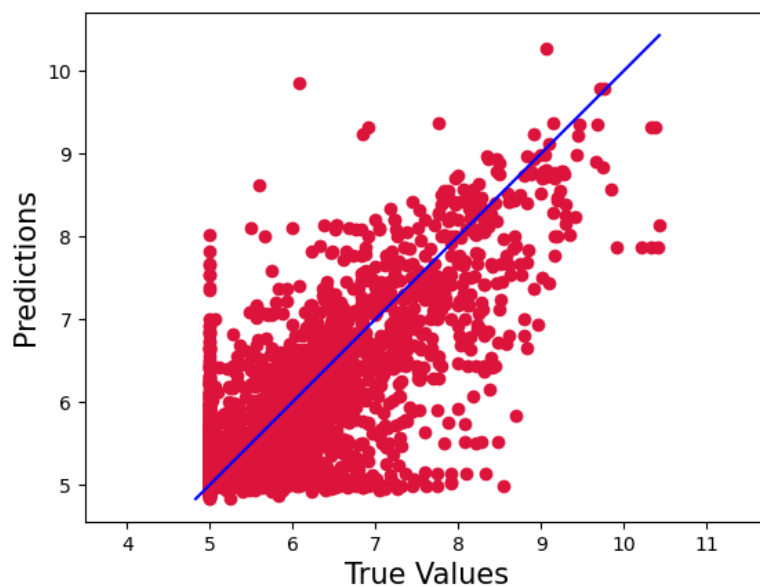


Figure 1. Predictions of the CSatDTA model vs. measured binding affinity values for the Davis dataset.

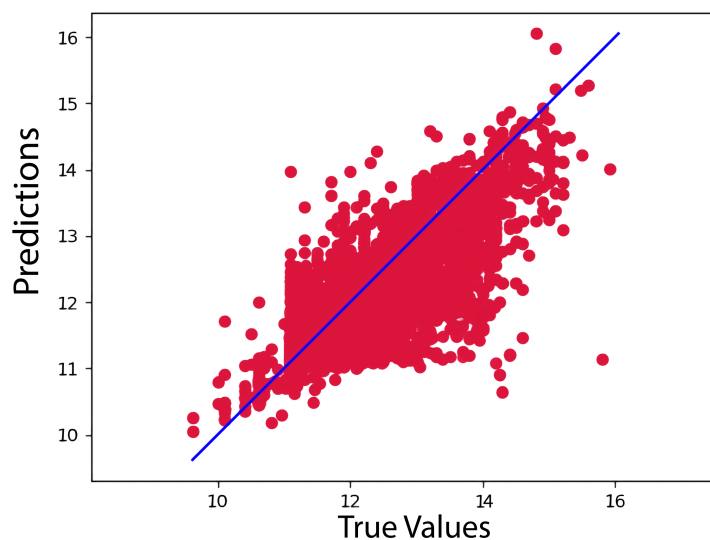


Figure 2. Predictions of the CSatDTA model vs. measured binding affinity values for KiBA.

Our findings show that attention enhancement leads to systematic gains in prediction and feature identification tasks across a wide range of architectures and computational methodologies. In addition, our experiments demonstrated the efficacy of the suggested 2D relative attention mechanism. We used self-attention feature maps instead of convolutional feature maps in all trials to allow straightforward comparisons with the baseline models. All outcomes were consistent with the 2D relative self-attention mechanism unless otherwise stated.

values are derived from various data sources, including IC_{50} , K_i , and K_d , whereas Davis is derived exclusively from K_d . The number of drugs, targets, and interactions, as well as the statistics of the datasets are provided in Table 4.

Table 4. Datasets.

	Proteins	Compounds	Interactions
KIBA	229	2111	118,254
Davis	442	68	30,056

4.2. Drug and Target Representation

For drugs, the input is fed as SMILES data, a form readable by computers [32]. The SMILES are parsable with a context-free parser. This representation has been used while predicting biochemical properties, including toxicity and biodegradability, based on the fundamental principle of cheminformatics, that similar molecules have similar properties [11]. With the SMILES code, drug properties such as the bulky atom number or amount of valence electrons can be retrieved and used as attraction prediction features. A string representation of the SMILES code can be found here. Natural language processing (NLP) techniques or a deep learning model can be used to highlight the strings. SMILES data for all compounds were collected from the PubChem database. After obtaining the SMILES string, it was passed through a unique label representation of the strings. In total, 64 labels were utilized. Each label is represented with corresponding integers. The data distribution from the Davis and KiBA datasets is shown in Figure 4.

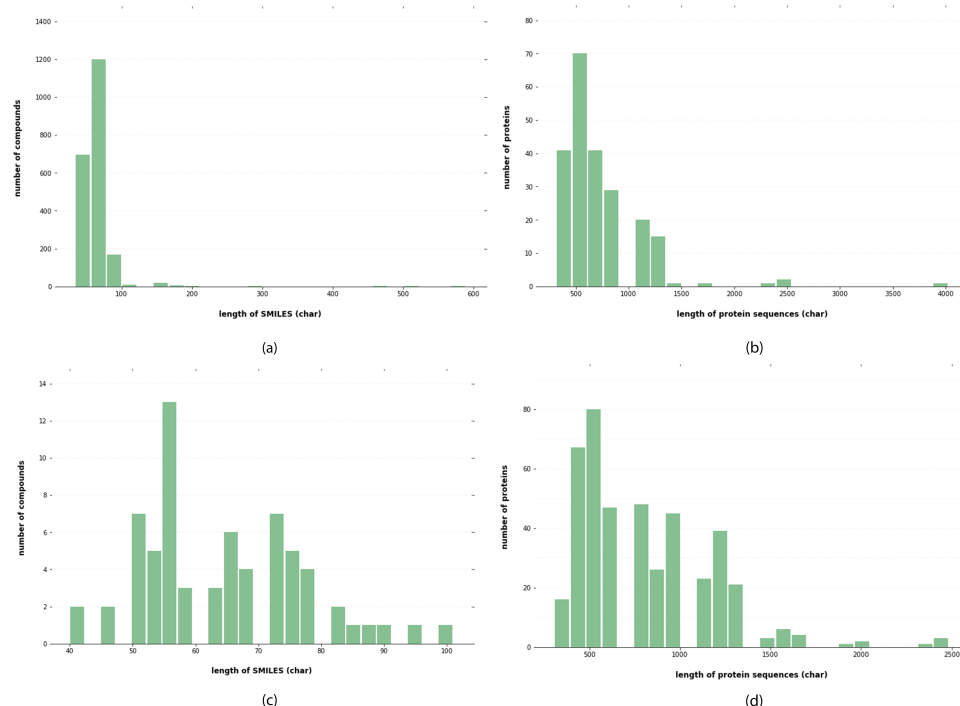


Figure 4. Analysis of KiBA and Davis datasets: (a) distribution of length of SMILES for KiBA datasets, (b) distribution of length of protein sequences for KiBA datasets, (c) distribution of length of SMILES for Davis datasets, and (d) distribution of length of protein sequences for Davis datasets.

Proteins are represented using one-hot encoding. For each target in the experimental datasets, the UniProt database was used to retrieve protein sequences with the help of the protein gene name. A series of ASCII characteristics representing amino acids constituted the sequence. Each amino acid type was assigned an integer based on its alphabetical symbol, and proteins are represented as an integer sequence. Aspartic acid (D) is four;

alanine (A) is one; cysteine (C) is three. The sequence was shortened or padded to a preset length of 1000 residues to make training easier. When a sequence was too short, zero values were inserted to extend it. Similarly, we determined a maximum character length of 100 for SMILES. The maximum length was determined by the distribution lengths of protein sequences and SMILES, as shown in Figure 4, to encompass at least 90% of the protein sequence and 80% of the SMILES compounds. Sequences shorter than the maximum limit were padded with zeros, whereas longer ones were terminated.

4.3. Proposed Prediction Model: CSatDTA

This research proposes CSatDTA, a self-attention augmented CNN-based model. The proposed model is motivated by the fact that target sequences and the structure of drugs are very comparable to natural language texts, in which atomic information, both structural and contextual, is crucial for comprehending the attributes of a molecule. The molecular structure of substances is encoded by SMILE sequences in which each atom interacts with both distant and nearby atoms. However, when representing a molecule, the current DeepDTA approach [21] employing CNNs is unable to link long-distance atoms. We used the self-attention technique to overcome this issue. Some studies have shown that combining convolutions with self-attention produces the optimum results [33]. In computer experiments, we found that the self-attention-aided convolution method outperforms convolutions as a stand-alone computational primitive in the DTA. We describe the CSatDTA model to enhance convolutional operators using the self-attention mechanism by concatenating feature maps of convolutions with a collection of feature maps created by self-attention. First, we discuss the input and output representations for the proposed CSatDTA model architecture. Following an explanation of the model training procedure, we go through the basic building blocks of the CSatDTA model.

As a regression problem, we computed the binding strength score from drug and target interaction forecasting in this work. As for estimation analytics, we chose a well-known deep neural network: a CNN with a self-attention technique, as shown in Figure 5. A CNN is a sort of architecture that consists of one or more convolutional layers, as well as a pooling layer. A max-pooling downsamples the previous layer's findings, allowing for the generalization of the characteristics acquired by the filters. The performance of the model was enhanced by an attention mechanism in addition to the CNN. A weighted average of the values computed using the concealed units was produced by self-attention. Our attention-augmented networks use self-attention throughout the design rather than pre-training, similar to their fully convolutional competitors. The multi-head attention (MHA) mechanism allows the model to give importance to both the spatial and feature subspaces at the same time. Furthermore, we extended the relative self-attention to 2D inputs, enabling us to model equivariance systematically, improving the representational capacity of self-attention. Instead of adding or reducing convolutional features, our method generates additional feature maps. This ability enables us to dynamically modify the proportion of attentional channels and evaluate a range of designs, from fully convolutional to attentional models.

Fully connected (FC) dense layers were added to the model in addition to the convolutional, attention, and max-pooling layers. The potential of convolutional networks to extract local dependencies using filters is their most significant feature. As a consequence, the size and number of filters in a CNN have a direct impact on the type of characteristics the model extracts from the feed. It is commonly assumed that as the number of filters in a model grows, so does the model's capacity to recognize patterns. We suggested a prediction model based on the attention CNN that consists of two self-attention-augmented convolutional blocks. In this study, d_k , d_v , and N_h relate to the depth of keys, the depth of values, and the number of heads in MHA, respectively. We also assume that d_k and d_v can be uniformly divided by N_h and denote d_k^h and d_v^h , respectively, as the depth of keys and values per attention head. Each convolutional block with self-attention augmentation seeks to learn representations from protein sequences/SMILES strings. We used an architecture of five convolution layers and one attention layer for each block.

The number of filters in the first two 1D convolutional layers increases with the number of layers, with the second layer having twice as many as the first. However, for the remaining convolution layers, the number of filters was manipulated with d_k and d_v in such a way that the final layer was reshaped to three-times the filter number in the first layer. A max-pooling layer then followed the neural network blocks. The max-pooling layers' final features were concatenated and then fed into the FC layers. In the first and second dense layers, we employed 1024 nodes each, which was accompanied by a 0.1 dropout layer. Dropout is a regularization strategy that involves turning off the activation of particular neurons to minimize overfitting [34]. The output layer was followed by the third layer, which had 512 nodes. Figure 5 depicts the proposed model, which combines two CNN blocks.

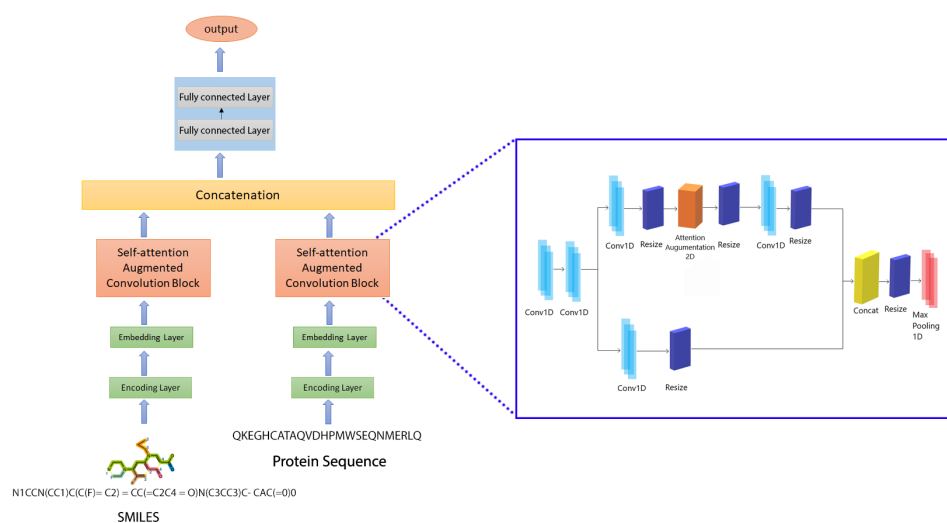


Figure 5. The proposed model's architecture.

4.3.1. Attention Mechanism

We used the Transformer architecture [35] to flatten an input tensor of shape (H, W, F_{in}) to a matrix $x \in R^{HW \times F_{in}}$ and conduct the MHA. For a single head h , the output of the self-attention mechanism can be described as:

$$O_h = \text{Softmax} \left(\frac{(XW_q)(XW_k)^T}{\sqrt{d_k^h}} \right) (XW_v).$$

where $W_v \in R^F$, W_q and $W_k \in R^{F_{in} \times d_k^h}$ are learned linear maps that convert the input X to queries $Q = XW_q$, values $V = XW_v$, and keys $K = XW_k$. After that, the outputs of all heads are combined and provided again as follows:

$$\text{MHA}(X) = \text{Concat}[O_1, \dots, O_{N_h}]W^o$$

where $W^o \in R^{d_v \times d_v}$ is a linear map that has been learned. To equal the pioneer dimensions, $\text{MHA}(X)$ is reshaped into a tensor of shape (H, W, d_v) .

4.3.2. Combining Attention and Convolutional Feature Mapping

Consider an original convolution operator with an F_{out} output filter, F_{in} input filters, and filter size k in formal terms. This corresponds to attention-enhanced convolution, which can be written as:

$$\text{AAConv}(X) = \text{Concat}[\text{Conv}(X), \text{MHA}(X)]$$

The ratio of key depth to the number of pioneer output filters is denoted by $k = \frac{d_k}{F_{out}}$, while the ratio of attentional channels to the number of pioneer output filters is denoted by $v = \frac{d_v}{F_{out}}$. The suggested attention-enhanced convolution is translation equivariant, like convolution, and may easily operate on inputs of many spatial dimensions.

4.3.3. Impact on the Number of Parameters

The MHA uses a linear convolution layer $(2d_k + d_v) = F_{out}(2k + v)$ output and F_{in} input kernels, as well as an added linear convolution layer with $d_v = F_{out}$ input and output kernels to mix the contributions of different heads for the computation of keys, values, and queries. After the number of filters in the convolutional section is reduced, the following changes occur in the parameters:

$$\Delta_{params} \sim F_{in}F_{out}(2k + (1 - k^2)v + \frac{F_{out}}{F_{in}}v^2)$$

For simplicity, the arguments introduced by relative position embeddings are ignored, benefiting from their negligibility. In practice, this results in a minor decrease in parameters when 3×3 convolutions are replaced and a slight increase in parameters for the replacement of 1×1 convolutions. Attention-augmented networks surpass fully convolutional networks by employing fewer test parameters, which is significant.

4.3.4. Attention Augmented Convolutional Architectures

In this analysis, we used boosted convolution with a batch normalization layer [36], which can adapt to adjust the convolution and attention feature mapping contributions. We utilized our boosted convolution per each residual block, similar to existing visual attention approaches [37,38]. In addition, we employed a limited batch size and often downsampled the inputs given to the self-attention layer. This lowers the memory consumption of attention-enhanced networks.

4.4. Evaluation Metrics

For evaluating the performance of these models, we applied the following metrics.

4.4.1. Concordance Index

The concordance index (CI) is considered as an analysis statistic to assure DTA predictive performance, as mentioned earlier [4]. The CI [17] is a ranking indicator for continuous data. The CI was used to measure how well the binding strength values of protein–ligand interactions are predicted with respect to the real values. The CI extends from $\frac{1}{2}$ to 1, with 1 denoting perfect prediction accuracy and $\frac{1}{2}$ denoting a random predictor. The following formula was used to calculate the CI:

$$CI = \frac{1}{Z} \sum_{\delta_i > \delta_j} h(b_i - b_j)$$

$s_i > s_j$, where b_i is the greater affinity's (δ_i) predicted value, whereas for the smaller affinity (δ_j), b_j is the prediction value. Here, (Z) is a normalization constant that equals the number of data pairings with different label values. The Heaviside step function is represented by $h(x)$, which is a discontinuous function defined as

$$h(x) = \begin{cases} 1.0, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0.0, & x < 0 \end{cases}$$

4.4.2. Mean-Squared Error

For continuous prediction errors, the MSE is a commonly used statistic parameter. The MSE, such as the variance, is expressed in the same units as the estimated quantity's

square. The root-mean-squared error (RMSE or RMSD) is obtained by taking the square root of the MSE, which has the same units as the predicted quantity. The RMSE equals the square root of the variance, often known as the standard error, for an unbiased estimator. Because this is a regression task, we picked the MSE as the statistic.

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

where p_i refers to the prediction, n to the number of samples, and y_i to the actual output

4.4.3. Root-Mean-Squared Error

The RMSE, which is among the linear regression measurements used in this study, is the mean distance between the predicted line and actual data. It is calculated as the square root of the MSE:

$$\text{RMSE} = \sqrt{\text{MSE}}$$

Author Contributions: All of the components of CSatDTA were co-developed by A.G. and H.T. The application interface was built by K.T.C. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant from the Korean government's National Research Foundation (NRF) (MSIT) (No. 2020R1A2C2005612).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this work can be downloaded from the project web page at <http://nscbio.jbnu.ac.kr/tools/CSatDTA/> (accessed on 15 July 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CSatDTA	Prediction of drug–target affinity using convolution model with self-attention
DTA	Drug–target affinity
CNN	Convolutional neural network
PDB	Protein Data Bank
MSE	Mean-squared error
CI	Concordance index
FC	Fully connected
MHA	Multi-head attention
SMILES	Simplified Molecular Input Line Entry System

References

- Gonzalez, M.W.; Kann, M.G. Chapter 4: Protein interactions and disease. *PLoS Comput. Biol.* **2012**, *8*, e1002819. [[CrossRef](#)] [[PubMed](#)]
- Weinshilboum, R. Inheritance and drug response. *N. Engl. J. Med.* **2003**, *348*, 529–537. [[CrossRef](#)] [[PubMed](#)]
- Mullard, A. New drugs cost US \$ 2.6 billion to develop. *Nat. Rev. Drug Discov.* **2014**, *13*, 877. [[CrossRef](#)]
- Ashburn, T.T.; Thor, K.B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683. [[CrossRef](#)]
- Lowe, M.M.; Blaser, D.A.; Cone, L.; Arcona, S.; Ko, J.; Sasane, R.; Wicks, P. Increasing patient involvement in drug development. *Value Health* **2016**, *19*, 869–878. [[CrossRef](#)] [[PubMed](#)]
- Chapagain, A.; Ghimire, A.; Joshi, A.; Jaiswal, A. Predicting Breast Cancer using Support Vector Machine Learning Algorithm. *Int. Res. J. Innov. Eng. Technol.* **2020**, *4*, 10.

7. Khanal, J.; Tayara, H.; Zou, Q.; Chong, K.T. Identifying dna n4-methylcytosine sites in the rosaceae genome with a deep learning model relying on distributed feature representation. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1612–1619. [[CrossRef](#)]
8. Kandel, J.; Tayara, H.; Chong, K.T. PUResNet: Prediction of protein–ligand binding sites using deep residual neural network. *J. Cheminformatics* **2021**, *13*, 1–14. [[CrossRef](#)]
9. Rehman, M.U.; Tayara, H.; Chong, K.T. DCNN-4mC: Densely connected neural network based N4-methylcytosine site prediction in multiple species. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 6009–6019. [[CrossRef](#)]
10. Rehman, M.U.; Akhtar, S.; Zakwan, M.; Mahmood, M.H. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images. *Biomed. Signal Process. Control* **2022**, *71*, 103212. [[CrossRef](#)]
11. Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* **2008**, *24*, i232–i240. [[CrossRef](#)] [[PubMed](#)]
12. Haggarty, S.J.; Koeller, K.M.; Wong, J.C.; Butcher, R.A.; Schreiber, S.L. Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem. Biol.* **2003**, *10*, 383–396. [[CrossRef](#)]
13. Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; Wei, Z. Drug–target affinity prediction using graph neural network and contact maps. *RSC Adv.* **2020**, *10*, 20701–20712. [[CrossRef](#)] [[PubMed](#)]
14. Gromiha, M.M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: Scoring schemes and binding affinity. *Curr. Opin. Struct. Biol.* **2017**, *44*, 31–38. [[CrossRef](#)] [[PubMed](#)]
15. Cer, R.Z.; Mudunuri, U.; Stephens, R.; Lebeda, F.J. IC 50-to-K i: A web-based tool for converting IC 50 to K i values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **2009**, *37*, W441–W445. [[CrossRef](#)]
16. Hu, P.W.; Chan, K.C.; You, Z.H. Large-scale prediction of drug–target interactions from deep representations. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 1236–1243.
17. Tian, K.; Shao, M.; Wang, Y.; Guan, J.; Zhou, S. Boosting compound-protein interaction prediction by deep learning. *Methods* **2016**, *110*, 64–72. [[CrossRef](#)]
18. Hamanaka, M.; Taneishi, K.; Iwata, H.; Ye, J.; Pei, J.; Hou, J.; Okuno, Y. CGBVS-DNN: Prediction of Compound-protein Interactions Based on Deep Learning. *Mol. Inform.* **2017**, *36*, 1600045. [[CrossRef](#)]
19. Pahikkala, T.; Airola, A.; Pietilä, S.; Shakyawar, S.; Szwajda, A.; Tang, J.; Aittokallio, T. Toward more realistic drug–target interaction predictions. *Briefings Bioinform.* **2015**, *16*, 325–337. [[CrossRef](#)]
20. Rose, P.W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Costanzo, L.D.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2016**, *45*, D271–D281.
21. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829. [[CrossRef](#)] [[PubMed](#)]
22. Ghimire, A.; Chapagain, A.; Bhattarai, U.; Jaiswal, A. Nepali Handwriting Recognition using Convolution Neural Network. *Int. Res. J. Innov. Eng. Technol.* **2020**, *4*, 5. [[CrossRef](#)]
23. Khanal, J.; Nazari, I.; Tayara, H.; Chong, K.T. 4mCCNN: Identification of N4-methylcytosine sites in prokaryotes using convolutional neural network. *IEEE Access* **2019**, *7*, 145455–145461. [[CrossRef](#)]
24. Abbas, Z.; Tayara, H.; Chong, K. Spinenet-6ma: A novel deep learning tool for predicting dna n6-methyladenine sites in genomes. *IEEE Access* **2020**, *8*, 201450–201457. [[CrossRef](#)]
25. Alam, W.; Ali, S.D.; Tayara, H.; Chong, K. A CNN-based RNA N6-methyladenosine site predictor for multiple species using heterogeneous features representation. *IEEE Access* **2020**, *8*, 138203–138209. [[CrossRef](#)]
26. Davis, M.I.; Hunt, J.P.; Herrgard, S.; Ciceri, P.; Wodicka, L.M.; Pallares, G.; Hocker, M.; Treiber, D.K.; Zarrinkar, P.P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046–1051. [[CrossRef](#)]
27. Tang, J.; Szwajda, A.; Shakyawar, S.; Xu, T.; Hintsanen, P.; Wennerberg, K.; Aittokallio, T. Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *J. Chem. Inf. Modeling* **2014**, *54*, 735–743. [[CrossRef](#)]
28. Nascimento, A.C.; Prudêncio, R.B.; Costa, I.G. A multiple kernel learning algorithm for drug–target interaction prediction. *BMC Bioinform.* **2016**, *17*, 1–16. [[CrossRef](#)]
29. 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. [[CrossRef](#)]
30. Öztürk, H.; Ozkirimli, E.; Özgür, A. WideDTA: Prediction of drug–target binding affinity. *arXiv* **2019**, arXiv:1902.04166.
31. Nguyen, T.; Le, H.; Quinn, T.P.; Nguyen, T.; Le, T.D.; Venkatesh, S. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37*, 1140–1147. [[CrossRef](#)]
32. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [[CrossRef](#)]
33. Bello, I.; Zoph, B.; Vaswani, A.; Shlens, J.; Le, Q.V. Attention augmented convolutional networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3286–3295.
34. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
36. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; PMLR: London, UK, 2015; pp. 448–456.
37. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-excite: Exploiting feature context in convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 9401–9411.
38. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.