CSConv2d: A 2-D Structural Convolution Neural Network with a Channel and Spatial Attention Mechanism for Protein-Ligand Binding Affinity Prediction

Xun Wang [1,2], **Dayan Liu** [1], **Jinfu Zhu** [3,*], **Alfonso Rodriguez-Paton** [4] **and Tao Song** [1,4,*]

[1] College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China; wangsyun@upc.edu.cn (X.W.); z19070030@s.upc.edu.cn (D.L.)
[2] High Performance Computer Research Center, Institute of Computer Technology, Chinese Academy of Sciences, Beijing 100190, China
[3] School of Economics, Beijing Technology and Business University, Beijing 100048, China
[4] Department of Artificial Intelligence, Faculty of Computer Science, Polytechnical University of Madrid, Campus de Montegancedo, 28660 Madrid, Spain; arpaton@fi.upm.es
[*] Correspondence: zhujinfuth@btbu.edu.cn (J.Z.); t.song@upm.es (T.S.)

Abstract: The binding affinity of small molecules to receptor proteins is essential to drug discovery and drug repositioning. Chemical methods are often time-consuming and costly, and models for calculating the binding affinity are imperative. In this study, we propose a novel deep learning method, namely CSConv2d, for protein-ligand interactions' prediction. The proposed method is improved by a DEEPScreen model using 2-D structural representations of compounds as input. Furthermore, a channel and spatial attention mechanism (CS) is added in feature abstractions. Data experiments conducted on ChEMBLv23 datasets show that CSConv2d performs better than the original DEEPScreen model in predicting protein-ligand binding affinity, as well as some state-of-the-art DTIs (drug-target interactions) prediction methods including DeepConv-DTI, CPI-Prediction, CPI-Prediction+CS, DeepGS and DeepGS+CS. In practice, the docking results of protein (PDB ID: 5ceo) and ligand (Chemical ID: 50D) and a series of kinase inhibitors are operated to verify the robustness.

Keywords: protein-ligand binding affinity; 2-D structural CNN; spatial attention mechanism

1. Introduction

Nowadays, there are a large number of drug candidate compounds in large databases such as ChEMBL [1] and PubChem [2]. The binding affinity of small molecules to receptor proteins is the key to drug discovery and drug repositioning [3]. Usually, chemical prediction methods are time-consuming and costly. Out of the 20,000 proteins in the human proteome, less than 3000 of them are targeted by known drugs. Simultaneously, in order to reduce the cost of drug development and the risk in the process of drug research, it has become an important strategy to repurpose the approved drugs and explore their new functions. Nevertheless, due to the huge amount of data and high-throughput screening (HTS), the discovery process of novel drugs is expensive and time-consuming [4].

The development of accurate prediction models for calculating binding affinity is imperative. Machine learning models are used to predict the interaction between drugs and targets on the basis of their chemical and biological characteristics, such as Support Vector Machine [5], Random Forest [6], deep learning methods [7–10] and so on. These major breakthroughs in deep learning have also pushed machine learning closer to the original human goal of "artificial intelligence". The main advantage of deep learning is that it can automatically learn data features using a common algorithm structure framework. This framework is usually composed of a stack of multilayer simple neural networks with nonlinear input and output mapping characteristics, including deep neural networks

(DNN), deep belief networks (DBN) [11], convolutional neural networks (CNN), autoencoders and recurrent neural networks (RNN).

Until now, plenty of deep learning frameworks and tools have been developed for various purposes in computational chemistry-based drug discovery, such as DeepDTA Drug-Target Affinity prediction [12]; GraphDTA using molecular graphs as the input of graph convolutional neural network [13]; and DeepPurpose integrating a variety of encoding methods of drug molecules and protein amino acid sequences for DTI prediction [14]. DeepGS inputs the sequence information and two-dimensional structure information of drug molecules as well as the protein sequence information into the model for prediction, while GraphDTA takes molecular graphs as the input of the model [15].

In this study, we propose a novel deep learning method for drug-target interactions (DTIs) prediction. The proposed model is named CSConv2d, which is improved from the original DEEPScreen model. It has a convolutional block attention module (CBAM) via the use of 2-D structural representations of compounds as the input instead of sequential features such as SMILES or molecular fingerprints [16]. Additionally, it is an added channel and spatial attention mechanism (CS) separate from the network structure. Functionally, the CBAM module can increase the nonlinear expression ability of the network and learn complex features from two-dimensional structure diagrams. Our CSConv2d can train a model for each target, and each model is independently optimized to accurately predict interacting small molecule ligands for a unique target protein. The main advantage of the module is increasing DTI prediction performances with the use of 2-D compound images instead of using conventional sequential features such as the molecular fingerprints and SMILES [17].

Data experiments are conducted on ChEMBLv23 datasets. The simulation results show that, in predicting protein-ligand binding affinity, our CSConv2d performs better than the original DEEPScreen model, as well as some state-of-the-art DTI prediction methods including DeepConv-DTI, CPI-Prediction, CPI-Prediction+CS, DeepGS and DeepGS+CS. In practice, the docking results of protein (PDB ID: 5ceo) and ligand (Chemical ID: 50D) are operated to verify the robustness.

2. Materials and Methods

2.1. Dataset

The ChEMBL database (v23) is employed to create the training dataset. ChEMBLv23 is a large, open-access drug discovery database that aims to capture Medicinal Chemistry data and knowledge across the pharmaceutical research and development process; it contains therapeutic targets and indications of clinical trial drugs and approved drugs, with a total of 1,961,462 different compounds and 13,382 targets.

In order to build a reliable training dataset, the data is filtered and preprocessed according to different types and bioactivity measurements (Figure 1). First, the data is filtered according to different attributes, such as "target type", "taxonomy", "assay type" and "standard type". Then, if multiple measurements for a ligand-receptor data point were present, the median value was chosen and duplicates were removed. For the "assay type", the functional assays were eliminated and only the binding assays were kept. Simultaneously, bioactivity measurements without the pCHEMBL value were removed; this value allows a number of roughly comparable measures of the half-maximal response concentration/potency/affinity to be compared on a negative logarithmic scale. After processing, the number of bioactivity points was 769,935. Bioactivities were extracted for activity values (IC50/EC50/Ki/Kd) of 10 μM or lower, with a CONFIDENCE_SCORE of 5 or greater for 'binding' or 'functional' human protein assays. For each target, compounds with bioactivity values $\leq$ 10 μM are selected as positive training samples and compounds with bioactivity values $\geq$ 20 μM are selected as negative samples. The 10 μM cutoff for activity specified here is in accordance with the method employed in the study of Koutsoukas et al. [18], representing both marginally and highly active compounds. The number of positive samples of each target is greater than the number of negative samples.

To balance it, a target similarity-based inactive dataset enrichment method is applied to populate the negative training sets to make the number of negative samples equal to the number of positive samples. The idea of this method is that similar targets have similar actives and inactives. The positive and negative sample ratio of the finalized training dataset for 800 target proteins is close to 1:1, which is 475,238 and 430,590, respectively.
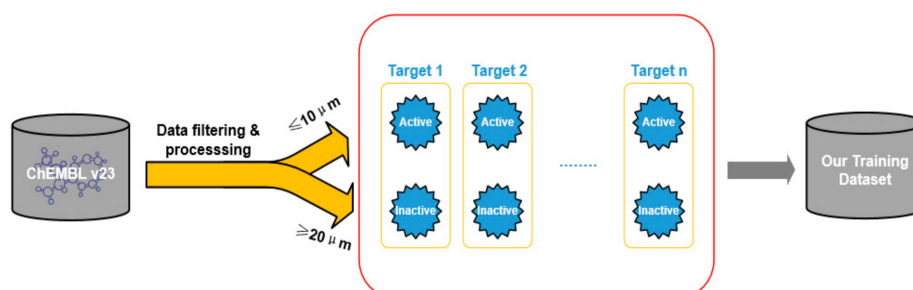

Figure 1. Data filtering and processing to create the training dataset of each target protein.

2.2. Attention Module

The CBAM is a lightweight and general module, which can be seamlessly integrated into any CNN architecture with negligible overheads and which is end-to-end trainable along with base CNNs. The module pay more attention to the object itself and thus has a better performance and interpretability. There are two independent submodules in CBAM, which are the channel attention module (CAM) and spatial attention module (SAM), as shown in Figure 2.

Figure 2. The overview of CBAM.

Since each feature map is equivalent to capturing a certain feature, channel attention helps to screen out meaningful features, that is, to "tell" the network which part of the original image features are meaningful. A pixel in the feature map represents a certain feature in an area of the original image, and spatial attention is equivalent to telling the network which area of the original image features should be paid attention to, ultimately improving the performance of the entire network.

2.3. Our CSConv2d Model

In general, our CSConv2d model is designed by embedding CBAM, to perform attention operations, in the framework of DEEPScreen [19]. The structure of our model is shown in Figure 3, which takes two-dimensional compound images as the input of the network.

In our model, each compound is represented by a 200-by-200-pixel 2-D image displaying the molecular structure generated by RDkit to make the representation unique and ensure the structure standard. Such a 2-D image is entered into the deep convolutional neural networks, which is composed of five convolutional + pooling, a channel attention module and a spatial attention module, and one fully connected layer preceding the output layer, and the DTI prediction was considered as a binary classification problem where the output could either be positive or negative. Each convolutional layer is followed by a ReLU activation function and max pooling layers. The last convolutional layer is flattened and connected to a fully connected layer, followed by the output layer. We use the Softmax

activation function in the output layer. The CSConv2d can train a model for each target, and each model is independently optimized to accurately predict interacting small molecule ligands for a unique target protein.
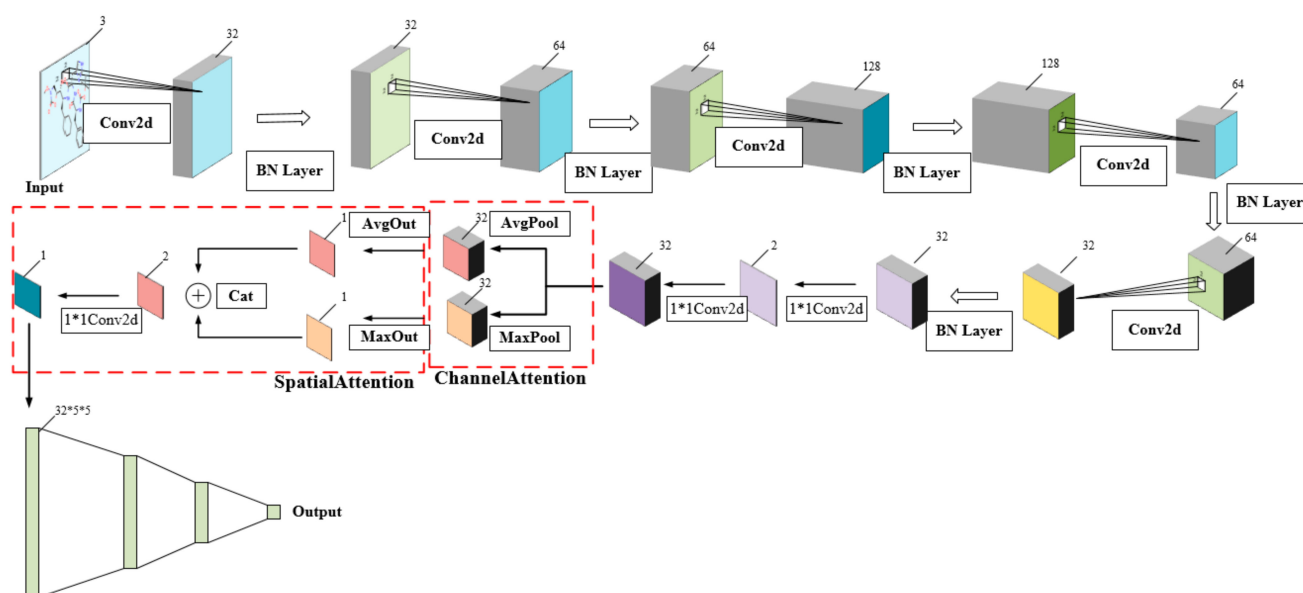

Figure 3. The structure of our model. It has five layers of convolution, channel attention and spatial attention blocks, and four layers of dense layer. CBAM blocks can not only effectively enhance the performance but are also computationally lightweight and impose only a slight increase in model complexity and computational burden.

2.4. Evaluation Metrics

We select the commonly used evaluation metrics, including Precision, Recall, Accuracy (ACC), F1-score and the Matthews correlation coefficient (MCC). The metrics used in the evaluation of our model are shown in Table 1.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \text{ Range } [0, 1] \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ Range } [0, 1] \tag{2}$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \text{ Range } [0, 1] \tag{3}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \text{ Range } [0, 1] \tag{4}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \text{ Range} [-1, 1] \tag{5}$$

Table 1. Comparison list of the mean values of accuracy (ACC), F1-score and Matthews correlation coefficient (MCC) between our model and DEEPScreen in a test set.

Model	ACC	F1-Score	MCC
CSConv2d (our)	0.83	0.84	0.67
DEEPScreen (Original)	0.81	0.81	0.63

TP represents the number of correctly predicted interacting drug-target pairs, and FN represents the number of interacting drug-target pairs that are predicted as noninteracting. TN represents the number of correctly predicted noninteracting drug-target pairs,

whereas FP represents the number of noninteracting drug-target pairs that are predicted as interacting.

3. Results

3.1. Performance of the CSConv2d Model

Data experiments are conducted on ChEMBLv23 datasets. The simulation results show that in predicting protein-ligand binding affinity, our CSConv2d performs better than the DEEPScreen model.

We trained and tested about 800 targets in total. The model is evaluated by ACC, F1-score and MCC. Each target is trained by a separated model, and their ACC, F1-score and MCC indexes are obtained respectively. The average value of the indicators of all the models is taken, as shown in Table 1. The performance of our CSConv2d is better than that of the original model. The ACC value of our model is improved by 2% and the value of the F1-score is improved by 3%. In addition, the MCC value calculated using Equation (5) is 0.67, demonstrating that the DTIs predicted by our model are more accurate than the original model.

We also compared the performance of the same datasets between our model and the original model using a scatter plot of different indexes (shown in Figure 4). As expected, the performance of our model was better than the original model.
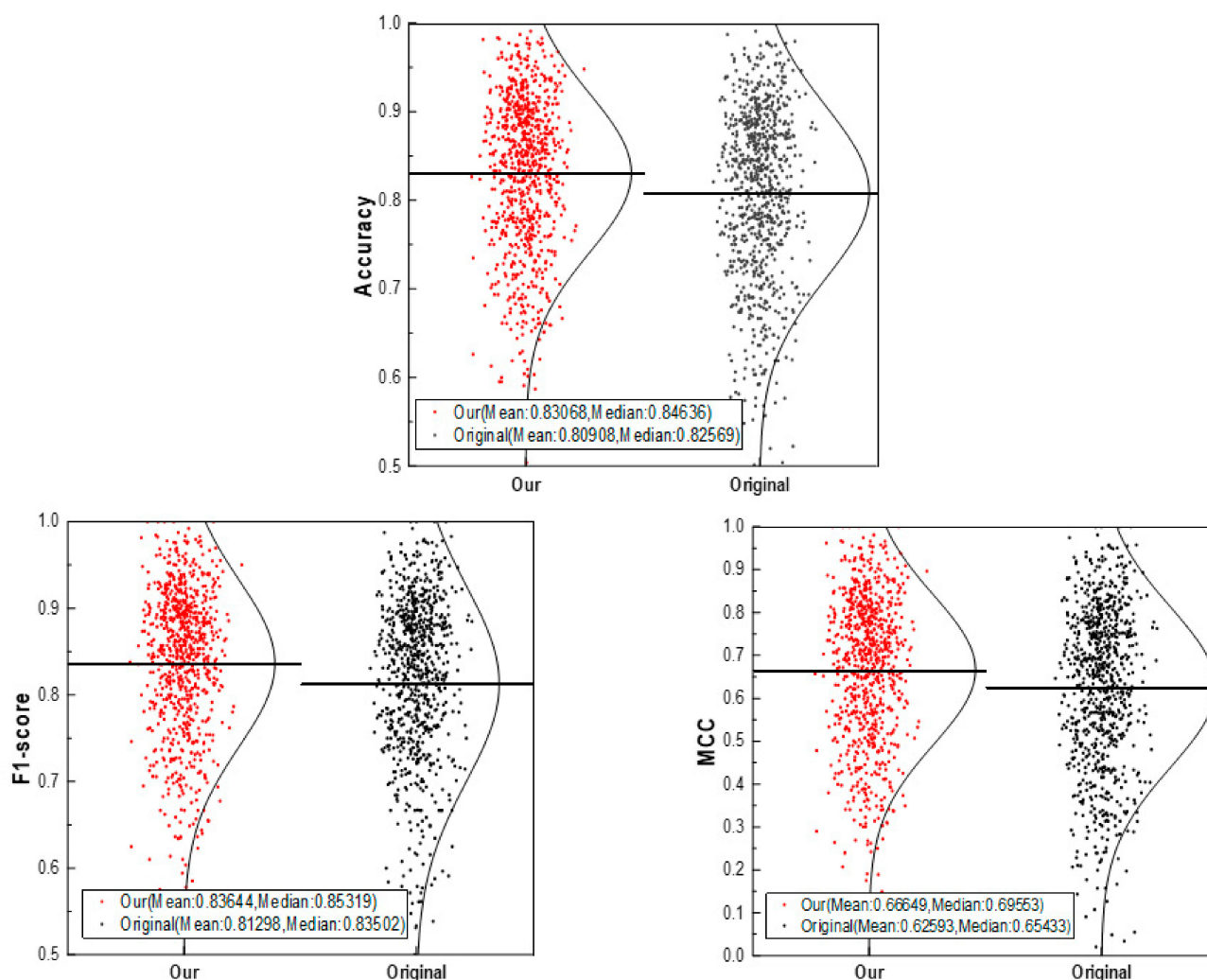

Figure 4. Performance comparison diagram of CSConv2d vs. original model.

3.2. Comparison with Different Models

We compare our CSConv2d with DeepConv-DTI proposed in 2019 [20], CPI-Prediction proposed in 2018 [21] and DeepGS proposed in 2020. The comparison results are shown in Table 2. In the comparison models, SMILES and the sequence of the protein were used as the input of the model. We randomly selected a target (Uniprot ID: P00797) trained in our model, then obtained the sequence of the protein and the SMILES of the compound dataset corresponding to the target as the input of the comparison models. In addition, we also added a channel attention mechanism and spatial attention mechanism to the structure of two of the models that were compared with CSConv2d. As shown in Table 2, it is found that the metrics of CSConv2d are higher than those of DeepConv-DTI, CPI-Prediction and DeepGS. In addition, in order to further reflect the performance of our model, we compared the models on a ChEMBL bioactivity benchmark set [22], and the results are shown in Table 3. On the benchmark set, the performance of our model is still the best.

Table 2. Comparison list of the accuracy (ACC), F1-score and Matthews correlation coefficient (MCC) of different models in a test set.

Model	ACC	F1-Score	MCC
DeepConv-DTI	0.79	0.80	0.64
CPI-Prediction	0.76	0.81	0.63
CPI-Prediction+CS	0.80	0.83	0.67
DeepGS	0.83	0.85	0.66
DeepGS+CS	0.85	0.86	0.70
CSConv2d	0.87	0.89	0.72

Table 3. Comparison list of the Matthews correlation coefficient (MCC) of different models in the ChEMBL Bioactivity Benchmark Set.

Dataset	Model	MCC
	Feed-forward DNN PCM	0.38
	CPI-Prediction+CS	0.42
ChEMBL Bioactivity Benchmark Set [23]	DeepGS+CS	0.47
	DEEPScreen	0.47
	CSConv2d	0.57

3.3. The Robustness

To investigate if CSConv2d could also improve the robustness of the original model, we randomly selected a protein, DLK in complex with inhibitor 2-((6-(3,3-difluoropyrrolidin-1-yl)-4-(1-(oxetan-3-yl)piperidin-4-yl)pyridin-2-yl)amino) isonicotinonitrile (PDB ID: 5ceo) [23], and its unique ligand's chemical ID was 50D from RCSB PDB [24], as shown in Figure 5. We first extracted the proto-ligand from the protein and redocked it using AutoDock Vina [25]. The RMSD calculated value between the redocked and RX-ligand is 1.3885. This is used as a benchmark. Furthermore, we randomly selected four compounds as a comparison. The results of the molecular docking show that all of the five compound molecules can dock with protein successfully, and the results are shown in Table 4. Simultaneously, interaction diagrams are provided in the Supplementary Materials (Figure S1). In particular, 50D, which is the original ligand of the protein, has the best performance of molecular docking.

According to the results of the prediction, the performance of our model is better than that of the DEEPScreen, which shows that our model has a better robustness. In addition, we also conducted comparative experiments on a series of kinase inhibitors, such as mTOR, VEGFR and JAK. The relevant data and results are shown in the Supplementary Materials (Table S1). The prediction performance of our model has been significantly improved.
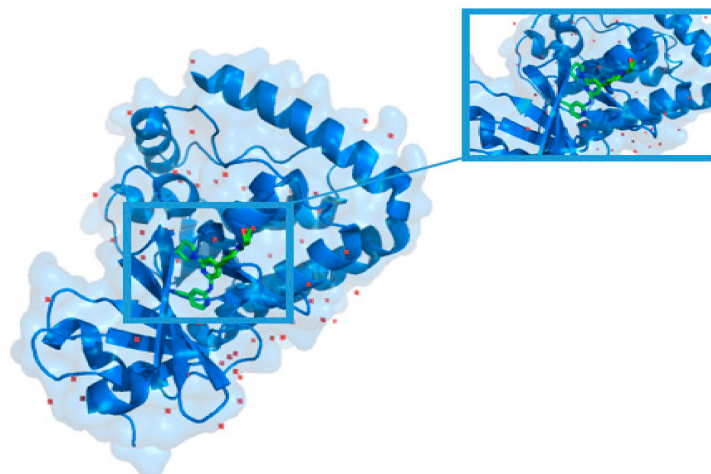
Figure 5. The complex of protein (PDB ID: 5ceo) and ligand (Chemical ID: 50D).

Table 4. Results of the molecular docking and the predicted result by CSConv2d and the original model.

Compound ID	Affinity (kcal/mol)	Pred (CSConv2d)	Pred (DEEPScreen)	Label
50D	−9.2	1	0	1
CHEMBL3731242	−8.7	1	1	1
CHEMBL3729274	−8.1	1	1	1
CHEMBL3355005	−7.9	1	1	1
CHEMBL3727745	−7.9	1	0	1
CHEMBL1242663	−7.5	0	0	0
CHEMBL1767275	−7.4	0	1	0
CHEMBL598911	−6.8	1	0	0
CHEMBL206659	−6.2	0	1	0
CHEMBL25829	−4.7	0	0	0

4. Conclusions

In this study, we propose a novel deep learning method for drug-target interactions (DTIs) prediction. The proposed model is named CSConv2d, which is improved from the original DEEPScreen model. It has a convolutional block attention module (CBAM) via the use of 2-D structural representations of compounds as input instead of sequential features such as SMILES or molecular fingerprint. Based on the CBAM, we added a channel attention mechanism and spatial attention mechanism after the last convolution layer of the network in order to improve the nonlinear expression of the network. The proposed CSConv2d can train a model for each target, and each model is independently optimized to accurately predict interacting small molecule ligands for a unique target protein. We also try to add them in the first layer of the network, but the effect is not good. We confirmed experimentally that CSConv2d outperformed the original model.

Data experiments are conducted on ChEMBLv23 datasets. The simulation results show that, in predicting protein-ligand binding affinity, our CSConv2d performs better than the original DEEPScreen model as well as some state-of-the-art DTI prediction methods including DeepConv-DTI, CPI-Prediction, CPI-Prediction+CS, DeepGS and DeepGS+CS. In practice, the docking results of protein (PDB ID: 5ceo) and ligand (Chemical ID: 50D) and a series of kinase inhibitors, such as mTOR, VEGFR and JAK, are operated to verify the robustness.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biom11050643/s1, Figure S1: The interaction diagram between the DLK in complex with inhibitor 2-((6-(3,3-difluoropyrrolidin-1-yl)-4-(1-(oxetan-3-yl)piperidin-4-yl)pyridin-2-yl)amino)

isonicotinonitrile (PDB ID: 5ceo) and 2-[[6-[3,3-bis(fluoranyl)pyrrolidin-1-yl]-4-[1-(oxetan-3-yl)piperidin-4-yl]pyridin-2-yl]amino]pyridine-4-carbonitrile (Chemical ID: 50D), Table S1: Results of molecular docking and predicted result by CSConv2d and DEEPScreen.

Author Contributions: Conceptualization, X.W. and T.S.; methodology, D.L.; software, D.L.; validation, J.Z., D.L. and A.R.-P.; formal analysis, A.R.-P.; investigation, X.W.; resources, T.S.; data curation, X.W.; writing—original draft preparation, T.S.; writing—review and editing, A.R.-P.; visualization, X.W.; supervision, X.W.; project administration, T.S.; funding acquisition, A.R.-P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw dataset is available for download at https://doi.org/10.4121/uuid:547e8014-d662-4852-9840-c1ef065d03ef (accessed on 31 July 2017).

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

References

1. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, 1100–1107. [CrossRef]
2. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, 1202–1213. [CrossRef]
3. Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner, C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijer, M.B.; Matos, R.C.; Tran, T.B.; et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181. [CrossRef]
4. Cao, Y.; Jiang, T.; Girke, T. A maximum common substructure-based algorithm for searching and predicting drug-like compounds. *Bioinformatics* **2008**, *24*, 366–374. [CrossRef]
5. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef] [PubMed]
6. Qi, Y. *Random Forest for Bioinformatics, Ensemble Machine Learning*; Springer: Boston, MA, USA, 2012; pp. 307–323.
7. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
8. Forsyth, D.A.; Ponce, J. *Computer Vision: A Modern Approach, 2/E[M]*; Prentice Hall: Hoboken, NJ, USA, 2002.
9. Halle, M.; Stevens, K. Speech recognition: A model and a program for research. *IEEE Trans. Inf. Theory* **1962**, *8*, 155–159. [CrossRef]
10. Chowdhury, G.G. Natural language processing. *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 51–89. [CrossRef]
11. Hinton, G.E. Deep belief networks. *Scholarpedia* **2009**, *4*, 5947. [CrossRef]
12. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34*, 821–829. [CrossRef] [PubMed]
13. Nguyen, T.; Le, H.; Quinn, T.P.; Venkatesh, S. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* **2020**, btaa921. [CrossRef] [PubMed]
14. Huang, K.; Fu, T.; Glass, L.M.; Zitnik, M.; Xiao, C.; Sun, J. DeepPurpose: A deep learning library for drug–target interaction prediction. *Bioinformartics* **2021**, *36*, 5545–5547. [CrossRef] [PubMed]
15. Lin, X.; Zhao, K.; Xiao, T. DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. *arXiv* **2020**, arXiv:2003.13902.
16. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
17. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Modeling* **1988**, *28*, 31–36. [CrossRef]
18. Koutsoukas, A.; Lowe, R.; KalantarMotamedi, Y.; Mussa, H.Y.; Klaffke, W.; Mitchell, J.B.; Glen, R.C.; Bender, A. In Silico target predictions: Defining a benchmarking data set and com-parison of performance of the multiclass Naïve Bayes and Parzen-Rosenblatt window. *J. Chem. Inf. Modeling* **2013**, *53*, 1957–1966. [CrossRef]
19. Rifaioglu, A.S.; Nalbat, E.; Atalay, V.; Martin, M.J.; Cetin-Atalay, R.; Doğan, T. DEEPScreen: High performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.* **2020**, *11*, 2531–2557. [CrossRef]
20. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129. [CrossRef]

21. Tsubaki, M.; Tomii, K.; Sese, J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* **2019**, *35*, 309–318. [CrossRef]

22. Lenselink, E.B.; Ten Dijke, N.; Bongers, B.; Papadatos, G.; Van Vlijmen, H.W.; Kowalczyk, W.; IJzerman, A.P.; Van Westen, G.J. Beyond the hype: Deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set (Dataset). 4TU.ResearchData. *J. Cheminformatics* **2017**, *9*, 1–14. [CrossRef]

23. Patel, S.; Harris, S.F.; Gibbons, P.; Deshmukh, G.; Gustafson, A.; Kellar, T.; Lin, H.; Liu, X.; Liu, Y.; Liu, Y.; et al. Scaffold-hopping and structure-based discovery of potent, selective, and brain penetrant N-(1 H-pyrazol-3-yl) pyridin-2-amine inhibitors of dual leucine zipper kinase (DLK, MAP3K12). *J. Med. Chem.* **2015**, *58*, 8182–8199. [CrossRef]

24. Rose, P.W.; Beran, B.; Bi, C.; Bluhm, W.F.; Dimitropoulos, D.; Goodsell, D.S.; Prlić, A.; Quesada, M.; Quinn, G.B.; Westbrook, J.D.; et al. The RCSB Protein Data Bank: Redesigned web site and web services. *Nucleic Acids Res.* **2010**, *39*, 392–401. [CrossRef] [PubMed]

25. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient op-timization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef] [PubMed]