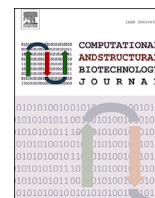




Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Mini-Review

Prediction of protein-ligand binding affinity with deep learning

Yuxiao Wang^{a,1}, Qihong Jiao^{a,1}, Jingxuan Wang^a, Xiaojun Cai^a, Wei Zhao^{b,*}, Xuefeng Cui^{a,*}^a School of Computer Science and Technology, Shandong University, Qingdao 266237, Shandong, China^b State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, Shandong, China

ARTICLE INFO

Keywords:

Protein-ligand affinity
 Binding affinity prediction
 Deep learning
 Assemble model

ABSTRACT

The prediction of binding affinities between target proteins and small molecule drugs is essential for speeding up the drug research and design process. To attain precise and effective affinity prediction, computer-aided methods are employed in the drug discovery pipeline. In the last decade, a variety of computational methods has been developed, with deep learning being the most commonly used approach. We have gathered several deep learning methods and classified them into convolutional neural networks (CNNs), graph neural networks (GNNs), and Transformers for analysis and discussion. Initially, we conducted an analysis of the different deep learning methods, focusing on their feature construction and model architecture. We discussed the advantages and disadvantages of each model. Subsequently, we conducted experiments using four deep learning methods on the PDBbind v.2016 core set. We evaluated their prediction capabilities in various affinity intervals and statistically and visually analyzed the samples of correct and incorrect predictions for each model. Through visual analysis, we attempted to combine the strengths of the four models to improve the Root Mean Square Error (RMSE) of predicted affinities by 1.6% (reducing the absolute value to 1.101) and the Pearson Correlation Coefficient (R) by 2.9% (increasing the absolute value to 0.894) compared to the current state-of-the-art method. Lastly, we discussed the challenges faced by current deep learning methods in affinity prediction and proposed potential solutions to address these issues.

1. Introduction

Proteins are responsible for a wide range of life activities in organisms, with up to three billion in human cells [1]. However, proteins cannot work alone in the body and must bind to other molecules, known as ligands [2,3]. These ligands interact with specific parts of proteins, known as protein pockets, to carry out various physiological functions [4]. During the binding process, the ligand is constantly changing shape to achieve the best fit in the protein pocket. Affinity is a measure used to assess the strength of the bond between the protein and ligand, with higher affinity indicating a stronger connection.

Protein-ligand interactions are essential for a variety of biological processes, such as interactions in antibody-antigen recognition [3], cell-cell communication [5], and signal transduction [6]. Abnormal interactions can lead to many diseases, so it is important to gain a thorough

understanding of the mechanisms behind these interactions to facilitate the development of new drugs [7]. However, the drug discovery process is complex and time-consuming [8], so it is necessary to develop efficient and accurate computational methods [9,10] to predict the affinity between proteins and ligands and accelerate drug discovery [11].

Traditional techniques such as GOLD [12], AutoDock [13], and X-Score [14] are commonly used to analyze protein-ligand interactions. However, these methods require specialized expertise in the relevant field and involve complex algorithms, making them time-consuming and difficult to implement. To address this issue, machine learning approaches have been proposed to predict the protein-ligand affinity, which significantly reduces manual operations that are complex and time-consuming. Li et al. [15] and Ballester et al. [16] used random forest models to predict the binding affinity of protein-ligand, while Nguyen et al. [17] applied the gradient boosting tree model

* Corresponding authors.

E-mail addresses: wei.zhao@sdu.edu.cn (W. Zhao), xfcui@email.sdu.edu.cn (X. Cui).¹ These authors contributed equally to the work.

<https://doi.org/10.1016/j.csbj.2023.11.009>

Received 24 August 2023; Received in revised form 3 November 2023; Accepted 3 November 2023

Available online 20 November 2023

2001-0370/© 2023 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Summary of deep learning algorithm-based methods.

Name	Feature	Model	Year
TopologyNet [30]	element-specific persistent homology (ESPH)	1D-CNN	2017
DeepSite [18]	3D voxel representation (16 Å ³)	3D-CNN	2017
K _{DEEP} [19]	3D voxel representation (24 Å ³)	3D-CNN	2018
Pafnucy [31]	3D voxel representation (20 Å ³)	3D-CNN	2018
DeepAtom [20]	3D voxel representation (32 Å ³)	3D-CNN	2019
Hu et al. [27]	protein sequence, molecular graph	Transformer, GAT	2020
GraphDTA [23]	protein sequence, molecular graph	1D-CNN, GNN	2020
Fusion [24]	3D voxel representation (48 Å ³), spatial graph representation	3D-CNN, GNN	2021
saCNN [21]	3D voxel representation (24 Å ³)	3D-CNN	2021
egGNN [25]	edge-gated graph feature	GNN	2021

and achieved excellent results. However, the process of this method is relatively complicated at first. The deep learning methods that have emerged recently are not only comparable to or even surpass this method in terms of results, but also show huge advantages in terms of time saving.

As deep learning advances in the areas of image recognition and natural language processing, researchers are increasingly exploring its application in the identification of potential drug candidates, and have achieved remarkable results. Jiménez et al. [18,19], Li et al. [20] and Wang et al. [21] have employed 3D Convolutional Neural Network [22] (3D-CNN) to predict protein-ligand binding affinities, which utilizes 3D structural information of both proteins and ligands. Nguyen et al. [23], Jones et al. [24] and Jiao et al. [25] have used Graph Neural Network [26] (GNN) to complete the affinity prediction of proteins and ligands, which takes into account the 3D structural information of proteins and structural information of ligands respectively. Hu et al. [27] have adapted the graph attention network [28] (GAT) model to extract the 2D structure information of proteins and ligands, and then applied the Transformer [29] model to further extract the sequence information of proteins, in an effort to incorporate more protein information.

We have gathered some approaches for predicting protein-ligand affinity, including traditional and deep learning methods. Deep learning methods are found to be more effective than traditional methods, so we will focus on the popular deep learning methods, which can be divided into three categories: convolutional neural network methods, graph neural network methods, and Transformer methods. We will first explain the feature construction techniques and model architectures used by various deep learning algorithms and their respective use cases. Then, we will compare the performance of these methods using the PDBbind benchmark dataset. Lastly, we will analyze four algorithms in detail and discuss their individual advantages in predicting protein-ligand affinity. Through these studies, we hope to provide useful information about the progress of deep learning methods and their effectiveness in addressing the issues related to protein-ligand affinity prediction.

2. Analysis of deep learning methods

At present, the most commonly used deep learning techniques for protein-ligand affinity prediction can be divided into three categories: convolutional neural networks, graph neural networks, and Transformers. Most of these models treat affinity as a continuous value and handle it as a regression task. The models we introduce next are all like this. In this section, we will look at the features of these three types of methods, examine their model building processes, and analyze their application scenarios. A summary of the related deep learning methods is presented in Table 1.

2.1. Methods based on convolutional neural network

Cang et al. [30] introduced the element-specific persistent homology (ESPH) method as a means of representing the three-dimensional spatial

structure of proteins and ligands using one-dimensional topological invariants. This approach effectively reduces computational complexity. Subsequently, they developed a multi-task and multi-channel topological neural network called TopologyNet, which utilizes fusion learning to enhance model performance. TopologyNet operates by converting 3D structure information into 1D representations, leading to some information loss. Furthermore, since the model employs a 1D convolutional neural network for feature learning, the overall model framework is simpler compared to a 3D network. However, compressing 3D structures into one dimension often leads to information loss, and then using a simple 1D convolutional neural network to learn features with incomplete information may make the results even worse. In fact, there are already many excellent models based on 3D spatial structure input and processed by 3D convolution. This also implies that we may get better results by directly using 3D neural networks to process spatial structures.

Jimenez et al. [18] first proposed to use 3D structure descriptors to construct the characteristics of proteins and ligands. Firstly, a 16 Å³ three-dimensional grid is constructed as a container to accommodate proteins and ligands. Then the relative coordinates of protein and ligand are calculated with the mean of ligand coordinates as the geometric central point. Then they only took the protein and ligand structures with relative coordinates in the three-dimensional grid (the whole ligand is basically in the grid) for feature construction. For proteins, the DeepSite mainly referred to the atomic type defined by AutoDock4 [32]. Then they fused the atomic information in the surrounding grids, and calculated the contribution of each atom through the following equation:

$$n(r) = 1 - e^{-(r_{vdw}/r)^{12}} \quad (1)$$

where r represents the Euclidean distance between the voxel of the current atom and other voxels, and r_{vdw} represents the van der Waals radius of the current atomic type. Once the features were constructed, they were inputted into a deep convolutional neural network (DCNN) for training. This method employs feature descriptors with a limited number of characteristic channels and a simple deep convolutional neural network to extract features from protein-ligand complexes and ultimately predict affinity.

Inspired by DeepSite, J Jimenez et al. proposed the K_{DEEP} model [19] again, as shown in Fig. 1(a). They adjusted DeepSite's descriptor set to better characterize proteins and ligands, including 8 characteristic channels: Hydrophobe, Aromatic, Acceptor, Donor, PosIonizable, NegIonizable, Metallic and Excluded volume. During the feature building process, they built a 24 Å³ three-dimensional grid instead of the 16 Å³, other feature building process is similar to DeepSite. Then they tried a variety of classical convolutional neural network architectures, including ResNet [33], VGG [34] and SqueezeNet [35]. Ultimately, they found that the architecture of the SqueezeNet model yielded the best results. By employing the improved feature representation and exploring various classical convolutional neural network architectures, they aimed to cover a wide range of possibilities. However, their investigations did

not lead to any significant improvement over the classical architectures tested.

The feature construction process of Stepniewska-Dziubinska et al. [31] was similar to that of DeepSite and K_{DEEP} . The key distinction in their approach was the utilization of a larger set of feature descriptors (19 descriptors) to represent the atoms in proteins and ligands. Additionally, they constructed a three-dimensional grid with a size of 20 \AA^3 . They defined a three-layer 3D convolutional neural network to learn features, and finally completed affinity prediction through three fully connected layers, called Pafnucy, as shown in Fig. 1(b). In summary, Pafnucy incorporates a more comprehensive range of protein and ligand information by employing additional feature descriptors, allowing the model to extract critical information more effectively from protein-ligand complexes.

Li et al. [20] introduced feature representations consisting of 12 descriptors to construct protein-ligand features. During the feature construction process, they employed a 32 \AA^3 three-dimensional grid to accommodate the protein-ligand complexes. The contribution of each atom in the protein and ligand was calculated using the same equation as DeepSite. Drawing inspiration from various lightweight network architectures [36–38], they proposed a novel lightweight 3D convolutional neural network. This network architecture improved the prediction performance without significantly increasing the model's complexity. These methods delve into the detailed exploration of the roles of various feature descriptors and convolutional neural networks in affinity prediction, thus advancing the progress of deep learning in predicting affinity.

Wang et al. [21] combined the K_{DEEP} feature representation method [19], and used the improved feature descriptors in HTMD [39] to construct features. To construct protein features, the researchers analyzed the physical and chemical properties of each atom and allocated them to specific channels accordingly. This process involved categorizing atoms based on their properties and assigning them to appropriate channels. For the feature construction of ligands, they employed the atomic types defined by SMARTS, which is a package used for specifying molecular patterns in cheminformatics. An open-source toolkit called feature factory used for constructing ligand features was implemented using RDKit [40] which can extract the chemical characteristics of each atom in the ligand. Throughout their experimentation, the researchers also explored different voxel sizes for constructing the three-dimensional grid. Ultimately, they determined that a grid size of 24 \AA^3 was most suitable for their purposes.

In Wang's work, in addition to enhancing the feature descriptor, a novel end-to-end convolutional neural network architecture called saCNN (spatial attention CNN) was proposed. This architecture incorporated a spatial attention mechanism, as depicted in Fig. 1(c). By applying the attention mechanism, the model could assign weights to different voxels, allowing it to focus more on important atom pairs or spatial structures. This, in turn, enabled the model to learn more profound features by prioritizing crucial information. The inspiration for the attention mechanism used in saCNN came from CBAM [41]. By integrating the attention mechanism into the 3D convolutional neural network, the saCNN model not only improved the existing feature descriptor but also facilitated easier learning within the model.

2.2. Methods based on graph neural network

Nguyen et al. [23] took the protein sequence as the text, and expressed the characteristics of the protein using the one-hot vector. At the same time, they transformed the SMILES code of the ligand into a molecular graph using RDKit software [40], and extracted five pieces of information as the feature representation of the ligand. Then they respectively constructed 1D convolution neural network and graph neural network models to learn the characteristics of proteins and ligands then finally connected the hidden layer characteristics of the two to achieve affinity prediction. When modeling ligands with graph neural network,

they tried four different models, including GCN [42], GAT [28], GIN [43], and GAT-GCN variants. They respectively regard ligands and proteins as two-dimensional molecular diagrams and one-dimensional sequences for feature learning. Although the SMILES, a data format used to represent molecular chemical formulas, and protein sequence data used for feature representation are easy to obtain, they lack the spatial structural information of protein-ligand complexes, which could lead to a decrease in model performance since protein-ligand interactions occur in a three-dimensional space. More and more work is tending to process and learn spatial structure, which also shows to a certain extent that spatial structure information does play an important role in affinity prediction.

Jones et al. [24] used 3D descriptors to represent proteins and constructed 3D spatial features. This process shared similarities with the convolution-based approach but focused solely on the 3D characterization of proteins. Ligands were treated as spatial molecular graphs, with atoms represented as nodes and covalent and non-covalent bonds as edges. To facilitate feature learning, both convolutional neural networks and graph neural networks were applied separately to process the protein and ligand data. The features of the protein and ligand were then fused to exchange information between them. The model makes full use of the 3D spatial structure information of protein and the 2D structure information of ligand, but the architecture of the model remained relatively simple, and there was no attention mechanism incorporated into the ligand modeling process, which made it challenging for the model to weigh the importance of each atomic node accurately.

Jiao et al. [25] characterized the inputs in the form of graph. The entire ligand was treated as a graph network, with atoms representing nodes and chemical bonds acting as edges connecting these nodes. The node features and edge features in graph networks were constructed by the RDKit tool. Node features were composed of various atom properties, including atom type, degree, chemical valence, aromaticity, formal charge, and free radical electrons. Edge features, on the other hand, encompassed bond type, aromaticity, conjugation, and ring information. In the case of proteins, a similar methodology was employed based on the practice of Torng et al. [44]. Each residue within the protein pocket was considered a node, and an edge was established between residues that were within a distance of 11 \AA . This procedure resulted in the construction of a protein pocket graph. The features of nodes and edges were from AAindex [45–47], which is an open source database containing various physicochemical properties of amino acids.

Then Jiao et al. [25] proposed an edge gated graph neural network model, called egGNN, viewed edges as gating units to control the flow of nodes in the graph, as shown in Fig. 1(d). The model integrated edge information in a novel way, which can learn the importance of different neighbor nodes (the importance of the same atoms connected by different chemical bonds are different). By utilizing the multi-head mechanism, the model achieved enhanced stability. Furthermore, the egGNN model employed the ReZero mechanism, enabling the training of deeper layers compared to traditional graph models. This mechanism facilitated the scalability of the model, ensuring its ability to handle complex datasets effectively. The egGNN not only uses more node features to describe proteins and ligands in the process of feature construction, but also proposes a new scalable graph neural network model framework to promote the weighted fusion of edges and nodes.

2.3. Methods based on transformer

When characterizing proteins, Hu et al. [27] applied both sequence information and two-dimensional structure information, which fully learned proteins to a certain extent. At the same time, they treated ligands as SMILES and also applied their two-dimensional structure information. The 2D feature representation of proteins and ligands reduces data sparsity and computational cost. Then they applied transformer

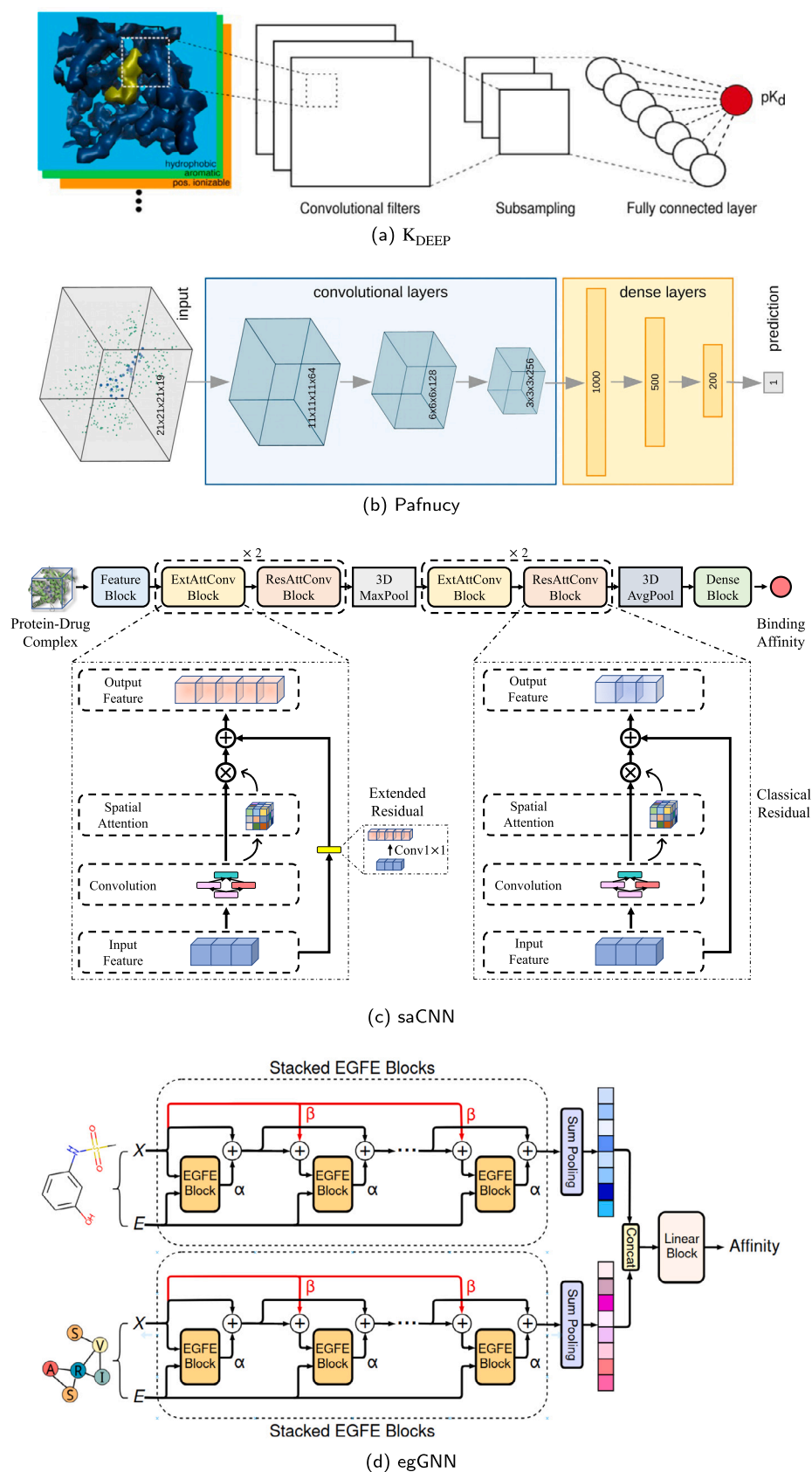
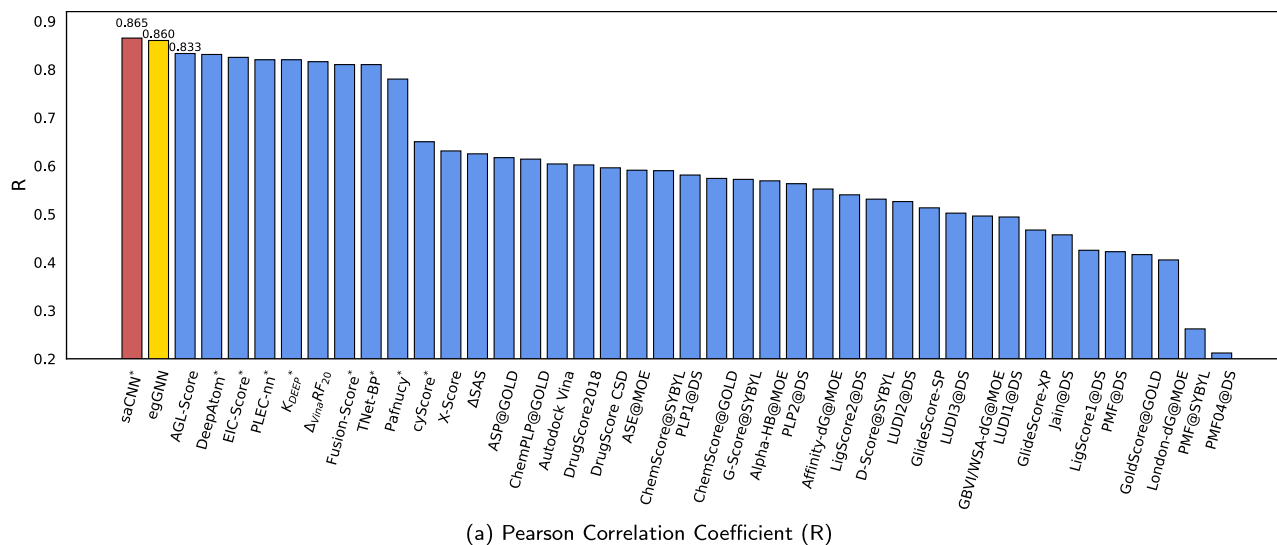
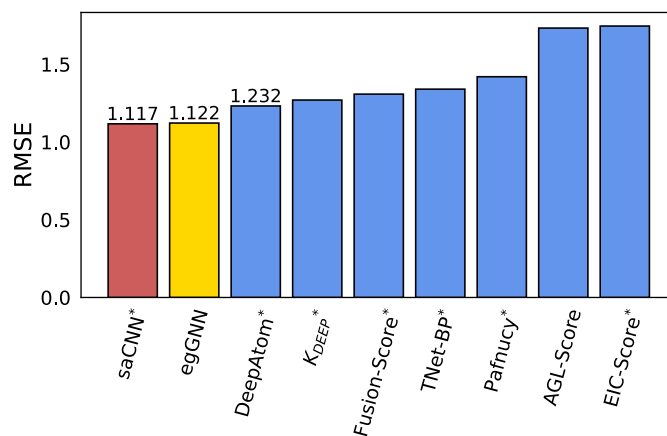


Fig. 1. The four deep learning models utilize either a Convolutional Neural Network (CNN) or a Graph Neural Network (GNN) to forecast the affinity of a protein-ligand pair. (a) Reproduced with the permission of ref. [19], © 2018 American Chemical Society. (b) Reproduced with the permission of ref. [31], © 2018 Oxford University Press. (c) and (d) Reproduced with the permission of ref. [21,25], © 2021 IEEE.



(a) Pearson Correlation Coefficient (R)



(b) Root Mean Square Error (RMSE)

Fig. 2. The results of saCNN (shown in red), egGNN (shown in yellow), and other methods (AGL-Score, K_{DEEP}, DeepAtom, CASF-2016, Fusion, TNet, PLEC, Pafnucy, EIC-Score, and cyScore) for predicting binding affinity are displayed for the PDBbind v.2016 core set (marked with *) and the CASF-2016 dataset (without *).

to pretrain protein sequences and applied GAT model to 2D structures of proteins and ligands to predict the affinity of protein-ligand. They explored the impact of different dimensional data of proteins on the model, and applied the attention mechanism in both proteins and ligands, making it easier for the model to find key information. However, there is still a lack of spatial structure information of proteins and ligands, and there is a certain degree of information loss.

Because protein sequences can be easily obtained from protein sequence databases, such as UniProt [48] which contains multiple sub databases, it is appropriate to apply Transformer model to proteins. However, Transformer training often requires a lot of computing resources and is time-consuming, and the use of protein sequence information will lose its spatial structure information, making the spatially adjacent atoms unable to be reflected in the sequence. There are few three-dimensional structure databases of proteins. However, due to the emergence of AlphaFold2 [49,50], the structure of proteins can be obtained accurately and quickly. Therefore, on the basis of a large number of protein structures, we can easily use 3D convolution neural network to model proteins, which can be faster than Transformer and retain the spatial structure information of proteins completely. After the 3D spatial structure information of the protein is obtained, the 2D structure information of the protein can also be obtained, so the graph neural

network can be used for modeling, which will be faster than the convolution neural network.

3. Results

This review focuses on four deep learning techniques: K_{DEEP} [19], Pafnucy [31], egGNN [25] and saCNN [21]. We conducted experiments to compare and contrast these methods in order to determine their respective strengths and weaknesses. This is essential for creating successful models for predicting protein-ligand binding affinity.

3.1. Datasets

In order to provide experimentally determined binding affinity data for all classes of biomolecular complexes stored in the Protein Data Bank [51] (PDB), the PDBbind database [52] was established. This database plays a critical role in bridging the gap between the energetic and structural information of these complexes, enabling various computational and statistical studies on molecular recognition, drug discovery, and related fields. PDBbind dataset provides an essential linkage between the energetic and structural information of those complexes, which is helpful for various computational and statistical studies on molecular

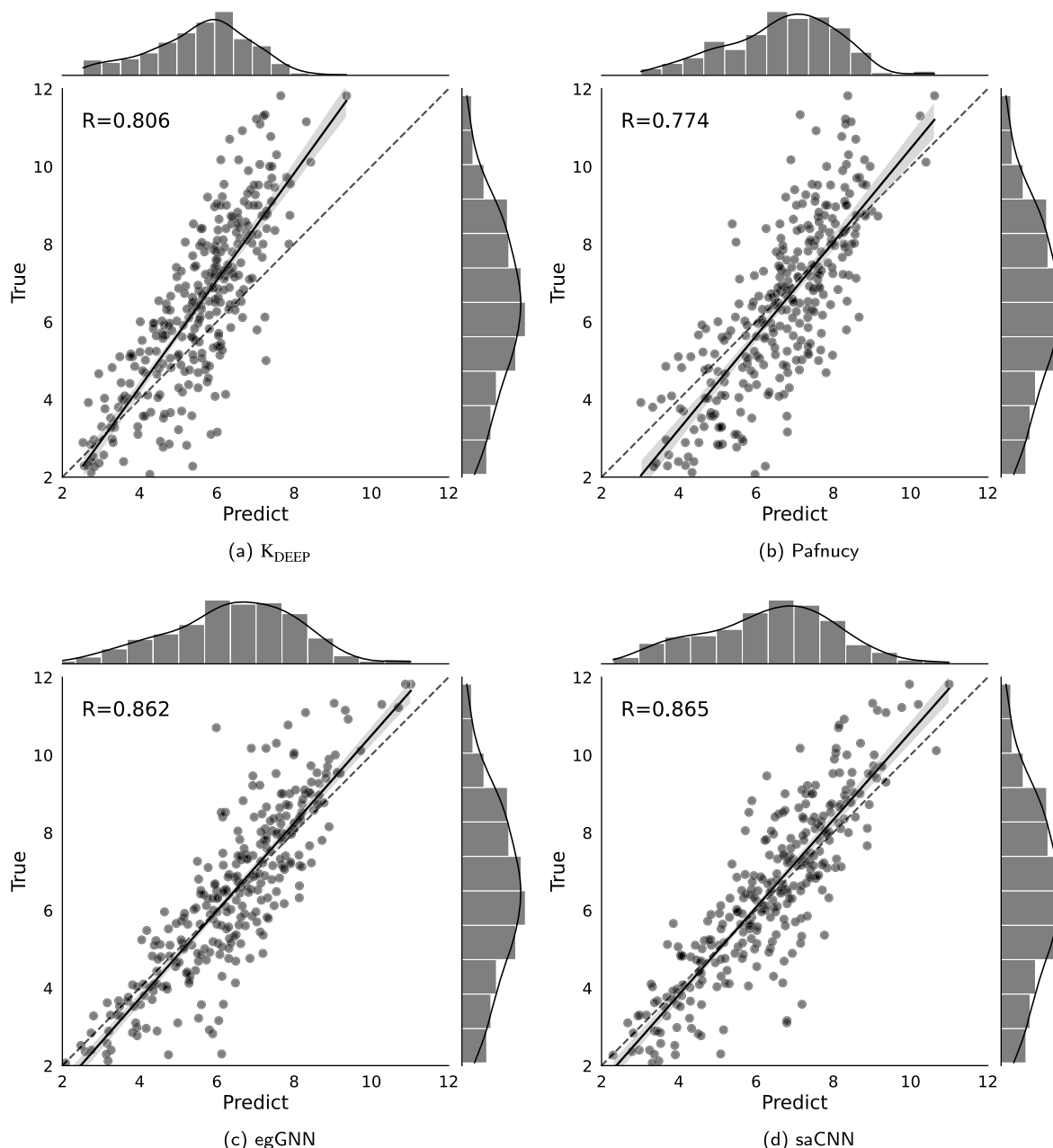


Fig. 3. The four models were compared by plotting the scatter plots between the experimental affinity and the predicted affinity of the PDBbind v.2016 core set. The experimental affinity was calculated using the formula $-\log(K_d/K_i)$, where K_d is the dissociation constant and K_i is the inhibition constant. The dotted line in the graph represents the ideal situation in which the predicted value is equal to the true value. The closer the points are to this line, the more accurate the model is in predicting the affinity.

recognition, drug discovery, and many more. According to the dissociation (K_d), inhibition (K_i), half-concentration (IC_{50}), and resolution factors, the PDBbind dataset can be divided into general set and refined set, which can be simply understood as normal quality and high quality. A total of 17342 protein-ligand complexes (excluding complexes that exist in the test set) from the general set and the refine set are used as the training set and 290 complexes are used as the test set for evaluation.

3.2. Performance and correlation analysis

Firstly, in order to explore the performance of various models in affinity prediction, we selected the Pearson Correlation Coefficient and the Root Mean Square Error (RMSE) between model predicted values and true affinity values (labels) as metrics. We listed some of the

currently representative methods. As shown in Fig. 2, the saCNN and egGNN methods have achieved top rankings in both of the metrics, securing the first and second positions. Meanwhile, K_{DEEP} and Pafnucy also demonstrated good performance in these two metrics. This not only indicates that K_{DEEP} and Pafnucy, both based on 3D convolution, perform well in affinity prediction tasks, but it also suggests that saCNN, utilizing spatial attention, and egGNN, utilizing graph neural networks, indeed further enhance the model's affinity prediction performance.

Secondly, to comprehensively compare the prediction performance of the four methods, correlation scatter plots were created for each method, as depicted in Fig. 3. The X-axis represents the predicted values generated by each method, while the Y-axis represents the true affinity values between the protein and ligand. The diagonal line indicates

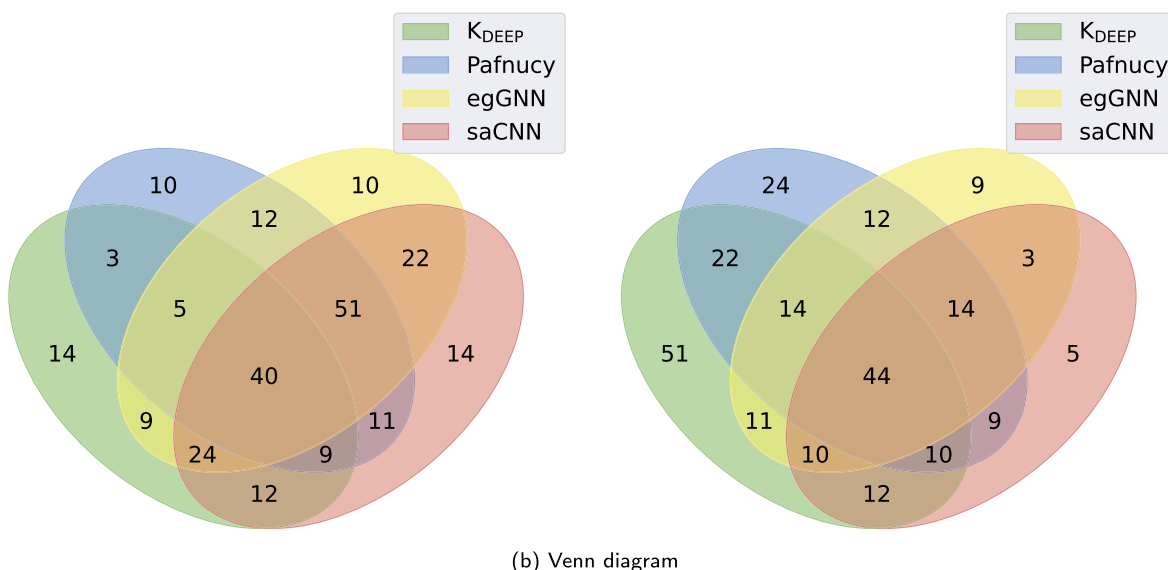
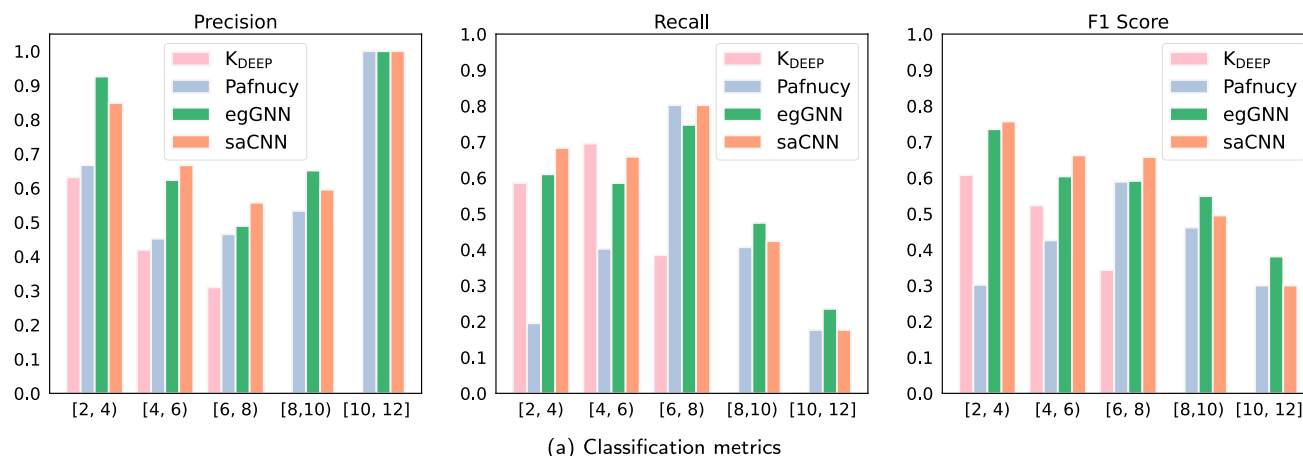


Fig. 4. The four models were evaluated in five distinct affinity ranges of proteins and ligands, and the results are presented. A Venn diagram is included to illustrate the true positives (left) and false positives (right) samples.

perfect prediction, where the predicted values align perfectly with the true values. Therefore, the closer the distribution of predicted and true values is to the diagonal, the more accurate the method's predictions. Upon comparison, K_{DEEP} exhibits significant deviations from the diagonal line, moreover, when the true affinity value is higher, the deviation becomes more pronounced, indicating poorer predictive performance. Pafnucy tends to follow the diagonal line more closely, although some deviations may occur when the true affinity value is relatively low. Additionally, from a distribution perspective, when the affinity value exceeds 10, the predicted values of K_{DEEP} and Pafnucy tend to be smaller. This discrepancy may be attributed to the limited availability of training data for this particular range, leading to incomplete learning by the models in this region. As for egGNN and saCNN, these two models perform relatively better compared to K_{DEEP} and Pafnucy, with the overall difference between them not being very significant. However, it's worth noting that saCNN exhibits a slight advantage when the true affinity value falls within the range of 4 to 8.

3.3. Analysis of different intervals

In order to explore the predictive ability of each method across different ranges of protein-ligand affinity, the affinity values were divided into five intervals ranging from 2 to 12. Precision, Recall, and F1 scores

were calculated for each method and plotted in Fig. 4 (a). From the figure, it can be observed that egGNN performs exceptionally well when the affinity between the protein and ligand exceeds 8. On the other hand, K_{DEEP} faces challenges in predicting affinities in these ranges, hence its absence in the corresponding sections of the figure. The saCNN method demonstrates superior performance when the affinity value falls within the range of 2 to 8. Furthermore, when the affinity value exceeds 10, the recall of all models is relatively low, indicating that very few methods accurately predict protein-ligand affinities above this threshold. However, their precision remains relatively high, suggesting that although most methods struggle to predict high-affinity interactions, the predictions made within this range are relatively reliable.

The true positive and false positive samples for each method were collected and represented using Venn diagrams in Fig. 4 (b). On the left side of the figure, it can be observed that the number of unique positive samples for each method is nearly identical. This suggests that each method possesses its own unique strengths and advantages, which may be attributed to the different strategies employed by each method. These advantages are particularly evident in certain individual samples. On the right side of the figure, it is evident that K_{DEEP} and Pafnucy exhibit a larger number of unique false positive samples compared to saCNN and egGNN. This indicates that spatial attention and graph neural networks have an advantage in extracting features from

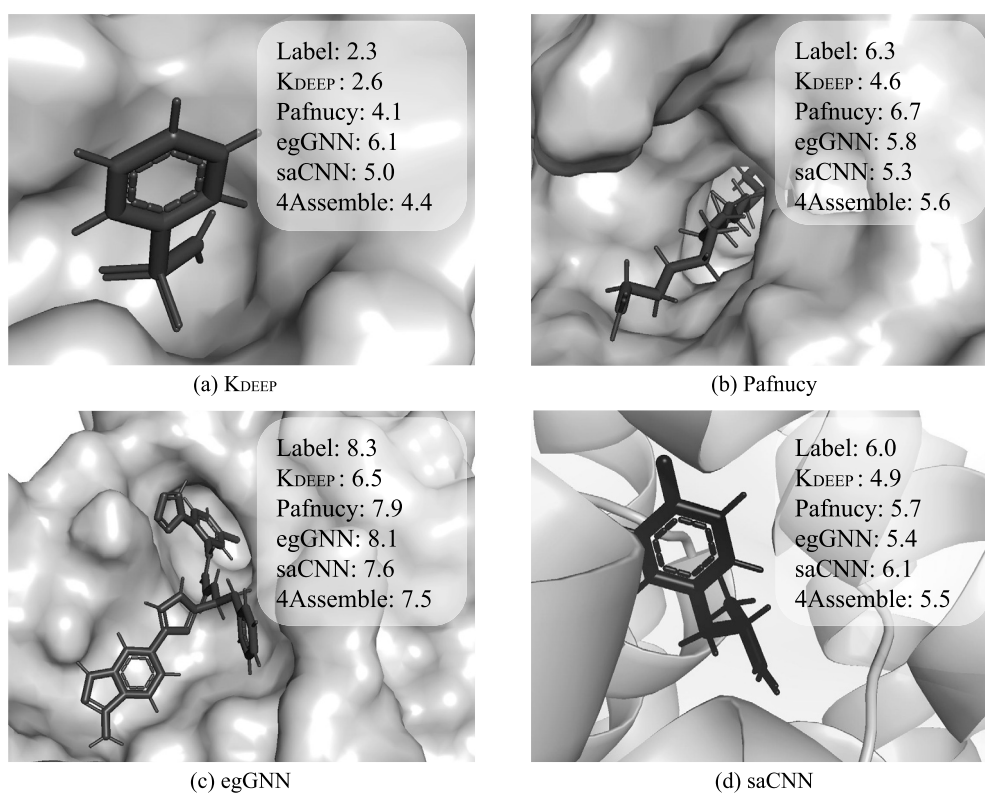


Fig. 5. The visualization of samples accurately predicted by each model is marked in the upper right corner with the true and predicted affinity values. The small molecules (in black) that each model excels at are distinct. K_{DEEP} is more accurate in predicting samples with smaller molecules. Pafnucy is better at predicting long strip molecules. egGNN is more successful when there are many rings in the molecule. saCNN is the most proficient in predicting small molecules covered by proteins.

Table 2

Comparison results of 4Assemble and the four methods.

Model	R	RMSE
saCNN	0.865	1.117
egGNN	0.862	1.121
K _{DEEP}	0.806	1.641
Pafnucy	0.774	1.424
4Assemble	0.894	1.101

protein-ligand complexes, leading to improved model performance and generalization.

Drawing inspiration from Fig. 4 (a), which highlights the unique advantages of each method, we hypothesized that aggregating these methods could lead to more reliable results across different types of samples. We consider these four models to exhibit outstanding performance in protein-ligand binding affinity prediction, and they each have distinct characteristics. For instance, K_{DEEP} and Pafnucy are both built upon 3D convolution, but the latter utilizes a larger set of feature descriptors (19 descriptors) to represent atoms. On the other hand, saCNN and egGNN are constructed based on spatial attention, and graph neural networks respectively. Their combination not only compensates for each other's shortcomings in feature extraction from protein-ligand complexes but also enhances the model's fault tolerance. To test this hypothesis, we integrated the four models into a combined model called 4Assemble, and performed experiments on the PDBbind v.2016 core set. The experimental results are presented in Table 2. The results demonstrate that the 4Assemble model achieves a correlation coefficient of 0.894 and an RMSE of 1.101, surpassing the performance of each individual method. Our hypothesis is that it is because the fusion of different models achieves complementary effects that the model has

stronger generalization capabilities. For example, 3D convolution cannot take into account the learning of global information very well, but spatial attention can supplement the problem of 3D convolution not being able to learn global information to a certain extent, at the same time, graph neural network can help learn the topological structure information of the data. To the best of our knowledge, these results outperform any existing method reported in the literature. These findings confirm our hypothesis that the integration of multiple models leverages the strengths of each method, resulting in improved prediction performance. It's worth emphasizing that we assign equal weight to each of the four models for the combination process.

3.4. Visualization

In order to gain a more concrete understanding of the differences between the methods, we conducted visualization experiments and presented the results in Fig. 5. We selected four representative samples that were better predicted by each method, providing insight into the binding states of proteins and ligands, as well as the structural characteristics of small molecules. Each example is selected based on the affinity value predicted by the four models, ensuring that each model has a best-performing example. In Fig. 5 (a), it can be observed that K_{DEEP} performs well on samples with smaller molecular structures. This suggests that K_{DEEP} may have a particular advantage in predicting affinity for such samples. Fig. 5 (b) reveals that Pafnucy excels in predicting the affinity of proteins and molecules with chain structures. This affinity prediction capability may be attributed to the specific features captured by Pafnucy that are relevant to this type of molecular configuration. Moving on to Fig. 5 (c), it is evident that egGNN demonstrates a superior performance on molecules with multiple rings. This observation is likely associated with the graph model employed by egGNN, which facilitates the effective integration of information from the molecular

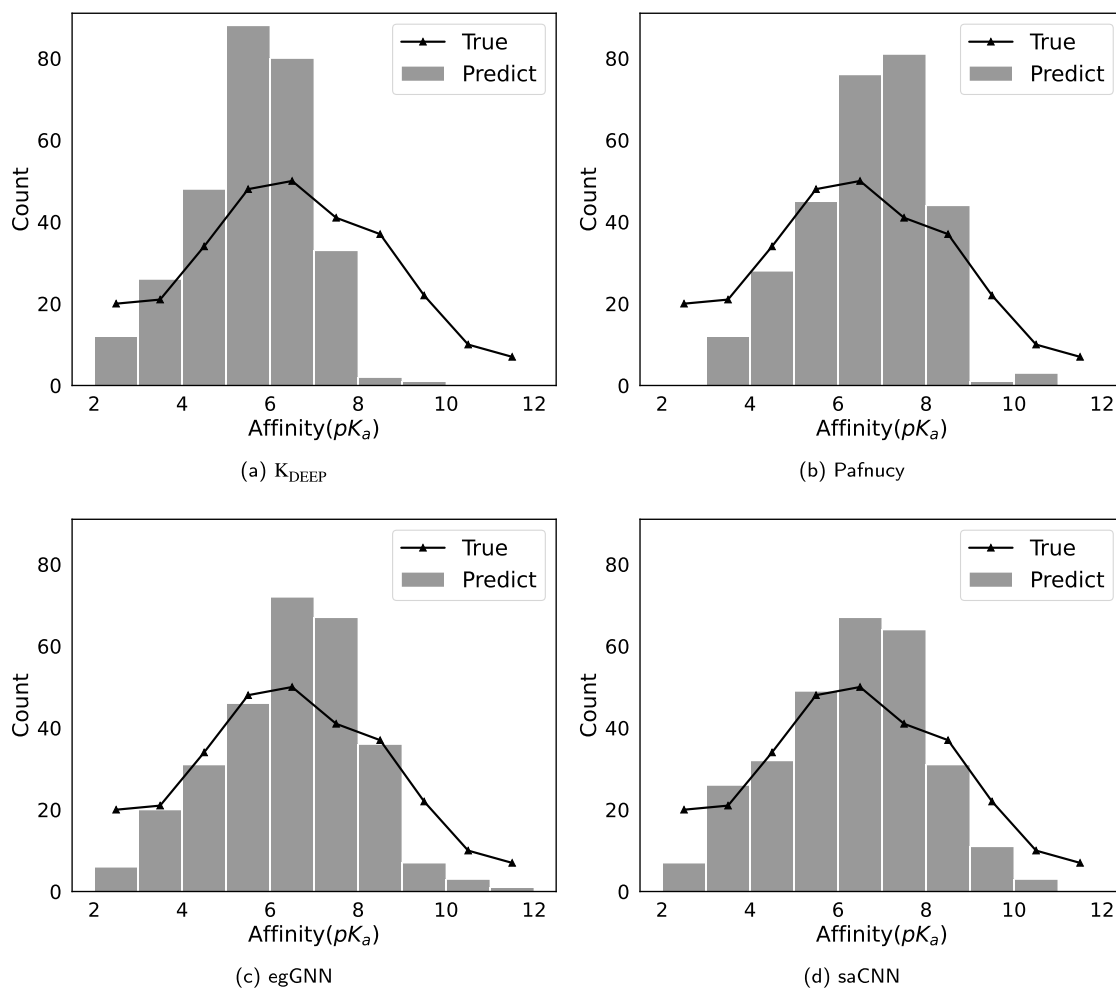


Fig. 6. The plots of the four models' experimental and predicted affinities on the PDBbind v.2016 core set show a commonality. This is because the training set has a lack of samples with both small and large affinity values, resulting in the models' poor predictions for those at either end.

graph, enabling it to capture relevant features specific to ring structures. Lastly, Fig. 5 (d) depicts a scenario where a molecule is enclosed within the protein. In this case, saCNN outperforms the other methods in accurately predicting the binding state. The advantage of saCNN can be attributed to its utilization of a 3D convolutional network with spatial attention, allowing it to effectively learn the spatial relationship and location information of the protein and ligand. It is worth noting that since Fig. 5 represents the best-case scenarios for K_{DEEP} , Pafnucy, saCNN, and egGNN, the 4Assemble model's performance may not be the absolute best. Nevertheless, its performance is acceptable and exhibits remarkable stability.

4. Discussions

Despite the progress made in predicting protein-ligand binding affinity, there are still some difficulties that need to be addressed. Fig. 6 shows the distribution of true and predicted values for each method, and it is clear that the performance of all four methods is weaker on samples with smaller or larger affinities. This is likely due to the lack of data in this range, which prevents the models from learning the relevant features for samples with extreme affinities. To tackle this issue, it is suggested to focus on this particular segment of the data during model training. Weighting or data augmentation can be used to improve the learning process for these samples, allowing the model to better capture the essential features of smaller or larger affinity values. Weighting or data augmentation has achieved great success in the field of CV. For example, AutoAugment [53] proposed by Google in 2019

is based on cropping, rotating, and translating images, and searching through the probability of corresponding operations to find the best augmentation strategy. Including the subsequently proposed Fast AutoAugment [54] and Population Based Augmentation [55], they have accelerated AutoAugment to a certain extent and achieved very good results. Even though there is currently no common weighting and augmentation method in the field of affinity prediction, we believe that similar methods will appear sooner or later, and we are also studying related weighting and augmentation methods. By addressing these challenges and following the advice given in this review, researchers can further enhance the performance of protein-ligand binding affinity prediction models.

5. Conclusion

Protein-ligand binding affinity prediction is a key factor in speeding up the drug development process. This paper examines the use of deep learning techniques for predicting protein-ligand affinity. Models based on convolutional neural networks are capable of capturing three-dimensional structural information of proteins and ligands, and the addition of attention mechanisms further improves their learning capabilities. On the other hand, models based on Transformers only learn the sequence information of proteins and ligands, thus losing valuable spatial information. Graph neural networks, however, are adept at capturing two-dimensional structural information by constructing simple graph features. Four deep learning methods were tested in this study, and it was observed that each model had distinct advantages

and disadvantages across different affinity intervals. An analysis of complexes predicted correctly or incorrectly by each model provided further insights into their prediction characteristics. Visualization of protein-ligand binding sites also helped to elucidate the strengths of each model. Finally, an ensemble model that combines predictions from all four models was developed through weighted integration, which showed improved binding affinity prediction capabilities. This integration approach leveraged the unique advantages of each individual model, leading to enhanced predictive performance. This study contributes to the understanding and application of deep learning methods for protein-ligand binding affinity prediction.

CRedit authorship contribution statement

Yuxiao Wang (First Author): Model design; Method integration; Most part of paper writing; Figure design; Experimental design.

Qihong Jiao (Second Author): Model design; Method integration; Most part of coding; Figure drawing; Experimental design.

Jingxuan Wang: Assistance in paper writing; Assistance in experimental design; Paper revision; idea Provision.

Xiaojun Cai: Assistance in model design; Assistance in experimental design; idea Provision.

Wei Zhao (Corresponding Author): Conceptualization, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.

Xuefeng Cui (Corresponding Author): Conceptualization, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.

Declaration of competing interest

All authors disclosed no relevant relationships.

Acknowledge on funding

This work was supported by the National Science Foundation of China (Grant No. 62072283) and the National Key R&D Program of China (Grant No. 2019YFA0905700, 2021YFC2101500).

References

[1] Milo Ron. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays, News Rev Mol Cell Dev Biol* Dec 2013;35(12):1050–5.

[2] Besson Benoit, Eun Hyeju, Kim Seonhee, Windisch Marc P, Bourhy Herve, Grailhe Regis. Optimization of BRET saturation assays for robust and sensitive cytosolic protein–protein interaction studies. *Sci Rep* Jun 2022;12(1):9987.

[3] Kobayashi Eiji, Jin Aishun, Hamana Hiroshi, Shitaoka Kiyomi, Tajiri Kazuto, Kusano Seisuke, et al. Rapid cloning of antigen-specific T-cell receptors by leveraging the cis activation of T cells. *Nat Biomed Eng* Apr 2022.

[4] Wu Canrong, Xu Youwei, He Qian, Li Dianrong, Duan Jia, Li Changyao, et al. Ligand-induced activation and g protein coupling of prostaglandin $\text{f}_2\alpha$ receptor. *Nat Commun* 2023;14(1):1–11.

[5] Armingol Erick, Officer Adam, Harismendy Olivier, Lewis Nathan E. Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* 2021;22(2):71–88.

[6] Sakaniwa Kentaro, Fujimura Akiko, Shibata Takuma, Shigematsu Hideki, Eki-moto Toru, Yamamoto Masaki, et al. Tlr3 forms a laterally aligned multimeric complex along double-stranded RNA for efficient signal transduction. *Nat Commun* 2023;14(1):164.

[7] Souza Paulo CT, Thallmair Sebastian, Conflitti Paolo, Ramirez-Palacios Carlos, Alessandri Riccardo, Raniolo Stefano, et al. Protein–ligand binding with the coarse-grained Martini model. *Nat Commun* 2020;11(1):1–11.

[8] Mizukoshi Yumiko, Takeuchi Koh, Tokunaga Yuji, Matsuo Hitomi, Imai Misaki, Fujisaki Miwa, et al. Targeting the cryptic sites: NMR-based strategy to improve protein druggability by controlling the conformational equilibrium. *Sci Adv* 2020;6(40):eabd0480.

[9] Li Li, Koh Ching Chiek, Reker Daniel, Brown JB, Wang Haishuai, Lee Nicholas Keone, et al. Predicting protein-ligand interactions based on bow-pharmacological space and Bayesian additive regression trees. *Sci Rep* 2019;9(1):7703.

[10] Souza Paulo CT, Thallmair Sebastian, Conflitti Paolo, Ramirez-Palacios Carlos, Alessandri Riccardo, Raniolo Stefano, et al. Protein-ligand binding with the coarse-grained Martini model. *Nat Commun* 2020;11(1):3714.

[11] Payandeh Jian, Volgraf Matthew. Ligand binding at the protein-lipid interface: strategic considerations for drug design. *Nat Rev Drug Discov* 2021;20(9):710–22.

[12] Jones Gareth, Willett Peter, Glen Robert C, Leach Andrew R, Taylor Robin. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997;267(3):727–48.

[13] Goodsell David S, Olson Arthur J. Automated docking of substrates to proteins by simulated annealing. *Proteins, Str Func Bioinform* 1990;8(3):195–202.

[14] Wang Renxiao, Lai Luhua, Wang Shaomeng. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput-Aided Mol Des* 2002;16(1):11–26.

[15] Li Hongjian, Leung Kwong-Sak, Wong Man-Hon, Ballester Pedro J. Low-quality structural and interaction data improves binding affinity prediction via random forest. *Molecules* 2015;20(6):10947–62.

[16] Ballester Pedro J, Mitchell John BO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* 2010;26(9):1169–75.

[17] Nguyen Duc Duy, Wei Guo-Wei. AGL-Score: algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. *J Chem Inf Model* 2019;59(7):3291–304.

[18] Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* 2017;33(19):3036–42.

[19] Jiménez José, Škalič Miha, Martínez-Rosell Gerard, De Fabritiis Gianni. K_{DEEP} : protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J Chem Inf Model* 2018;58(2):287–96.

[20] Li Yanjun, Rezaei Mohammad A, Li Chenglong, Li Xiaolin. DeepAtom: a framework for protein-ligand binding affinity prediction. In: 2019 IEEE international conference on bioinformatics and biomedicine (BIBM); 2019. p. 303–10.

[21] Wang Yuxiao, Qiu Zongzhao, Jiao Qihong, Chen Cheng, Meng Zhaoxu, Cui Xuefeng. Structure-based protein-drug affinity prediction with spatial attention mechanisms. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM); 2021. p. 92–7.

[22] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86(11):2278–324.

[23] Nguyen Thin, Le Hang, Quinn Thomas P, Nguyen Tri, Le Thuc Duy, Venkatesh Svetha. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021;37(8):1140–7.

[24] Jones Derek, Kim Hyojin, Zhang Xiaohua, Zemla Adam, Stevenson Garrett, Bennett WF Drew, et al. Improved protein–ligand binding affinity prediction with structure-based deep fusion inference. *J Chem Inf Model* 2021;61(4):1583–92.

[25] Jiao Qihong, Qiu Zongzhao, Wang Yuxiao, Chen Cheng, Yang Zhenghe, Cui Xuefeng. Edge-gated graph neural network for predicting protein-ligand binding affinities. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM); 2021. p. 334–9.

[26] Scarselli Franco, Gori Marco, Tsoi Ah Chung, Hagenbuchner Markus, Monfardini Gabriele. The graph neural network model. *IEEE Trans Neural Netw* 2008;20(1):61–80.

[27] Hu Fan, Hu Yishen, Zhang Jianye, Wang Dongqi, Yin Peng. Structure enhanced protein-drug interaction prediction using transformer and graph embedding. In: 2020 IEEE international conference on bioinformatics and biomedicine (BIBM); 2020. p. 1010–4.

[28] Veličković Petar, Cucurull Guillem, Casanova Arantxa, Romero Adriana, Liò Pietro, Bengio Yoshua. Graph attention networks. In: International conference on learning representations; 2018.

[29] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, et al. Attention is all you need. In: Advances in neural information processing systems; 2017. p. 5998–6008.

[30] Cang Zixuan, Wei Guo-Wei. TopologyNet: topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS Comput Biol* 2017;13(7):1–27.

[31] Stepniewska-Dziubinska Marta M, Zielenkiewicz Piotr, Siedlecki Pawel. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* 2018;34(21):3666–74.

[32] Morris Garrett M, Huey Ruth, Lindstrom William, Sanner Michel F, Belew Richard K, Goodsell David S, et al. Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;30(16):2785–91.

[33] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

[34] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. Available from: arXiv:1409.1556, 2014.

[35] Iandola Forrest N, Moskewicz Matthew W, Ashraf Khalid, Han Song, Dally William J, Keutzer Kurt. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1 MB model size. Available from: arXiv:1602.07360, 2016.

[36] Ma Ningning, Zhang Xiangyu, Zheng Hai-Tao, Sun Jian. ShuffleNet V2: practical guidelines for efficient CNN architecture design. In: Ferrari Vittorio, Hebert Martial, Sminchisescu Cristian, Weiss Yair, editors. Computer vision – ECCV 2018. Cham: Springer International Publishing; 2018. p. 122–38.

- [37] Sandler Mark, Howard Andrew, Zhu Menglong, Zhmoginov Andrey, Chen Liang-Chieh. Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4510–20.
- [38] Huang Gao, Liu Shichen, Van der Maaten Laurens, Weinberger Kilian Q. Condensenet: an efficient DenseNet using learned group convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 2752–61.
- [39] Doerr S, Harvey MJ, Noé Frank, De Fabritiis G. HTMD: high-throughput molecular dynamics for molecular discovery. *J Chem Theory Comput* 2016;12(4):1845–52.
- [40] Landrum. RDKit: open-source cheminformatics. Release 2014.03.1; 2010.
- [41] Woo Sanghyun, Park Jongchan, Lee Joon-Young, Kweon In So. CBAM: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 3–19.
- [42] Kipf Thomas N, Welling Max. Semi-supervised classification with graph convolutional networks. Available from: arXiv:1609.02907, 2016.
- [43] Xu Keyulu, Hu Weihua, Leskovec Jure, Jegelka Stefanie. How powerful are graph neural networks? Available from: arXiv:1810.00826, 2018.
- [44] Wen Torng, Altman Russ B. Graph convolutional neural networks for predicting drug-target interactions. *J Chem Inf Model* 2019;59(10):4131–49.
- [45] Tomii Kentaro, Kanehisa Minoru. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein Eng Des Sel* 1996;9(1):27–36.
- [46] Kawashima Shuichi, Ogata Hiroyuki, Kanehisa Minoru. AAindex: amino acid index database. *Nucleic Acids Res* 1999;27(1):368–9.
- [47] Kawashima Shuichi, Kanehisa Minoru. AAindex: amino acid index database. *Nucleic Acids Res* 2000;28(1):374.
- [48] Apweiler Rolf, Bairoch Amos, Wu Cathy H, Barker Winona C, Boeckmann Brigitte, Ferro Serenella, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2004;32:D115–9.
- [49] Jumper John, Evans Richard, Pritzel Alexander, Green Tim, Figurnov Michael, Ronneberger Olaf, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9.
- [50] Evans Richard, O'Neill Michael, Pritzel Alexander, Antropova Natasha, Senior Andrew, Green Tim, et al. Protein complex prediction with alphafold-multimer. bioRxiv; 2021.
- [51] Berman Helen M, Westbrook John, Feng Zukang, Gilliland Gary, Bhat Talapady N, Weissig Helge, et al. The protein data bank. *Nucleic Acids Res* 2000;28(1):235–42.
- [52] Wang Renxiao, Fang Xueliang, Lu Yipin, Wang Shaomeng. The PDBbind database: collection of binding affinities for protein - ligand complexes with known three-dimensional structures. *J Med Chem* 2004;47(12):2977–80.
- [53] Cubuk Ekin D, Zoph Barret, Mane Dandelion, Vasudevan Vijay, Le Quoc V. Autoaugment: learning augmentation policies from data. Available from: arXiv:1805.09501, 2018.
- [54] Lim Sungbin, Kim Ildoo, Kim Taesup, Kim Chiheon, Kim Sungwoong. Fast autoaugment. *Adv Neural Inf Process Syst* 2019;32.
- [55] Ho Daniel, Liang Eric, Chen Xi, Stoica Ion, Abbeel Pieter. Population based augmentation: efficient learning of augmentation policy schedules. In: International conference on machine learning. PMLR; 2019. p. 2731–41.