



## OPEN ACCESS

## EDITED BY

Leyi Wei,  
Shandong University, China

## REVIEWED BY

Jiazhou Chen,  
South China University of Technology,  
China  
Yushan Qiu,  
Shenzhen University, China

## \*CORRESPONDENCE

Yue Zhang,  
zhangyue@gpnu.edu.cn

## SPECIALTY SECTION

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 31 August 2022

ACCEPTED 30 September 2022

PUBLISHED 14 October 2022

## CITATION

Zhang Y, Hu Y, Li H and Liu X (2022),  
Drug-protein interaction prediction via  
variational autoencoders and  
attention mechanisms.  
*Front. Genet.* 13:1032779.  
doi: 10.3389/fgene.2022.1032779

## COPYRIGHT

© 2022 Zhang, Hu, Li and Liu. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Drug-protein interaction prediction *via* variational autoencoders and attention mechanisms

Yue Zhang\*, Yuqing Hu, Huihui Li and Xiaoyong Liu

School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China

During the process of drug discovery, exploring drug-protein interactions (DPIs) is a key step. With the rapid development of biological data, computer-aided methods are much faster than biological experiments. Deep learning methods have become popular and are mainly used to extract the characteristics of drugs and proteins for further DPIs prediction. Since the prediction of DPIs through machine learning cannot fully extract effective features, in our work, we propose a deep learning framework that uses variational autoencoders and attention mechanisms; it utilizes convolutional neural networks (CNNs) to obtain local features and attention mechanisms to obtain important information about drugs and proteins, which is very important for predicting DPIs. Compared with some machine learning methods on the C.elegans and human datasets, our approach provides a better effect. On the BindingDB dataset, its accuracy (ACC) and area under the curve (AUC) reach 0.862 and 0.913, respectively. To verify the robustness of the model, multiclass classification tasks are performed on Davis and KIBA datasets, and the ACC values reach 0.850 and 0.841, respectively, thus further demonstrating the effectiveness of the model.

## KEYWORDS

drug-protein interactions (DPIs), variational autoencoder (VAE), attention mechanism, convolutional neural network (CNN), deep learning - artificial neural network

## Introduction

Finding gene-drug relationships is important not only for understanding a certain mechanism of drug molecules, but also for developing treatments for patients. The gene-drug relationship is many-to-many, which is much more complex than a gene-to-drug or a drug-to-gene, and also explains the complex relationship between gene-drug. The gene-drug relationship has similarities to the drug-protein relationship (Chen et al., 2019; Huang et al., 2021).

In the prediction of RNA-binding proteins, limited by the huge cost of biological experiments, it is difficult to fully understand the underlying mechanisms of alternative splicing (AS) and related RNA-binding proteins (RBPS) in regulating the epithelial-mesenchymal transition (EMT) process. This needs to be achieved by means of

computational methods (Qiu et al., 2021b) proposed an inductive matrix-based model to study the relationship between RBP and AS during EMT. The main purpose of the model is to compensate for missing and unknown RBP-AS relationships (Qiu et al., 2021a) proposed a method based on weighted data fusion with sparse matrix tri-factorization to conduct experiments. The AS-RBP relationship is explored by assigning different weights to the source data. Both methods achieve good results. At the same time, this has parallels with the drug-protein relationship. It achieves the desired effect by looking for a drug to inhibit a binding site of a protein.

Drug-protein interactions (DPIs) exploration is a critical step in the drug discovery process. With the discovery of new drugs, the field of drug development continues to expand, and awareness regarding the repositioning of existing drugs and new interactions involving approved drugs is of increasing concern (Oprea and Mestres, 2012). Based on biological experiments, it usually takes 10–20 years and much money (US\$ 0.5–260 million) to develop a new drug (Avorn, 2015), so it is important to explore the interactions between drugs and proteins. In recent years, computer-aided methods have achieved good results and contributed significantly to the prediction of DPIs. The application of artificial intelligence in chemical research can accelerate the development of high-precision DPIs prediction methods.

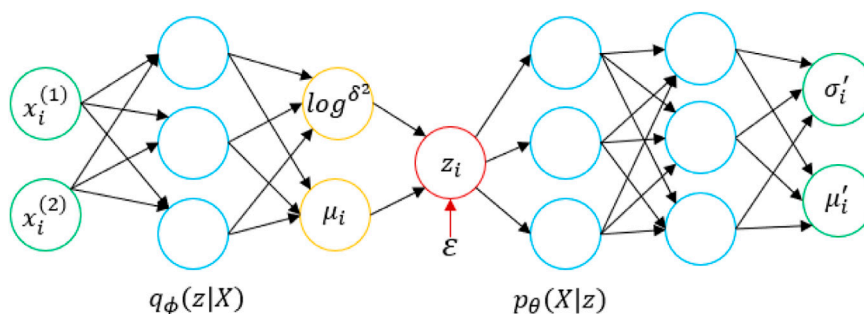
In the past decade, the problem of predicting the interactions between drugs and proteins has been solved using traditional machine learning methods, which solve binary classification problems (Yamanishi et al., 2010; Liu et al., 2016; Nascimento et al., 2016; Keum and Nam, 2017). Due to the rise and popularity of deep learning, it has become a popular choice for solving DPIs predictions (Unterthiner and Mayr, 2014; Tian et al., 2016) used a deep neural network (DNN) to explore the interactions between drugs and proteins instead of traditional machine learning methods, which directed the subsequent research on drug and protein interactions toward deep learning approaches, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) (Gao et al., 2018; Mayr et al., 2018) and stacked autoencoders (Wang et al., 2018).

In general, DPIs approaches can be divided into three categories: docking-based methods, machine learning-based methods, and deep learning-based methods. Docking-based methods require the best site and protein structure to be found and combined, but such a technique usually time-consuming, and many datasets lack three-dimensional protein structures (Gschwend et al., 1996). Machine learning-based methods (Faulon et al., 2008; Bleakley and Yamanishi, 2009; Ballester and Mitchell, 2010) usually require manual features, and the features passed to the model before modeling occurs require manual participation, which demands considerable feature extraction experience and expertise. Deep learning-based methods have been applied to many fields in biology (Min et al., 2016; Zeng et al., 2019; Zhang et al., 2019, 2021;

Wu et al., 2022b); DPIs prediction performance has been improved through the framework structure and network parameters of deep learning. For example, the DeepDTA approach of (Öztürk et al., 2018) learns internal high-level features by extracting the features of drugs and proteins as the network inputs and then predicts the relationships between drugs and proteins. The WideDTA method proposed by (Öztürk et al., 2019) is similar to DeepDTA, and the network framework is roughly unchanged; the main difference is that when inputting features, WideDTA extracts the features of drug proteins from multiple aspects as model inputs. Notably, a graph-based network architecture called GraphDTA (Nguyen et al., 2019), which treats drugs as a graph structure to predict DPIs, has also been developed. A Novel Graph Neural Network for Predicting Drug-Protein Interactions called BridgeDTA (Wu et al., 2022a), which introduces a class of nodes named hyper-nodes, which bridge different proteins/drugs to work as the protein-protein and drug-drug associations. HOGMMNC is a higher order graph matching with multiple network constraints model. It mainly obtains the fixed structural relationship in multi-source data through hypergraph matching, so as to identify the relationship between genes and drugs, and improve the accuracy and reliability of the identification relationship (Chen et al., 2019). These deep learning methods all have three similarities. 1) They encode drugs and proteins. 2) They extract the high-level features of drugs and proteins through their network structures. 3) They predict the features obtained in 2) through a fully connected (FC) layer. The advantage of these methods is that the process is not too cumbersome (it is simple). Furthermore, we exploit the strengths of these network frameworks for the prediction of DPIs.

A variational autoencoder (VAE) is a machine learning model that can reconstruct a variable  $x$  based on a latent feature  $Z$ . Unlike a simple autoencoder, it can learn the distribution of latent variables and then sample from this distribution to generate new samples. The model has been shaped and used in various fields, such as image processing (Walker et al., 2017; Liu et al., 2018) and text processing (Mayr et al., 2018). We use this model to predict DPIs. Experimental results show that better results are achieved on some datasets. Our contributions are as follows.

- 1) A variational autoencoder is designed to provide a probabilistic way of describing the latent representation of drugs and proteins, denoted *via* mean and variance of the hidden state distribution. Such generative way effectively reduce the redundant information in the raw samples to ease for learning drug-protein interactions.
- 2) Discriminative local features on drugs and proteins are extracted *via* deep CNN. A specially designed attention mechanism is incorporated to focus on the key interactive information on both drugs and proteins, thus obtaining strong drug-to-drug and protein-to-protein relationships.



**FIGURE 1**  
Structural diagram of the utilized VAE.

3) Extensive experiments on C.elegans and Human dataset, BindingDB dataset, Davis dataset and KIBA dataset. Datasets demonstrate that the proposed method can robustly identify the drug-protein interactions.

## Methods

### A VAE network to identify drug and protein interactions

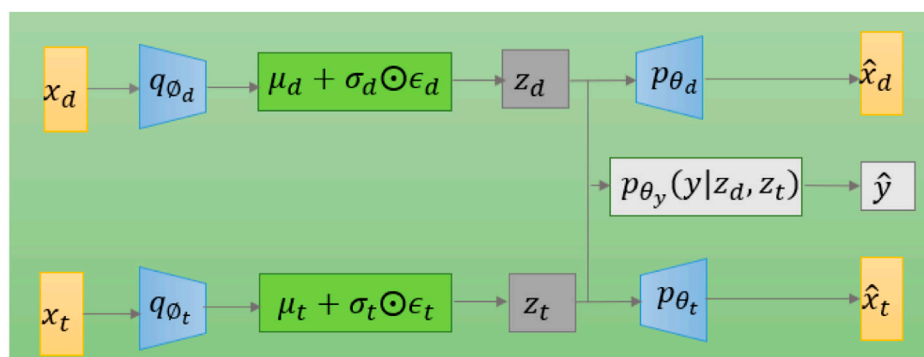
We input a set of drug molecules  $D$  and a protein sequence  $T$ , and a VAE (Kingma and Welling, 2014) learns the distribution of a multidimensional variable  $x$  based on an independent and identically distributed latent variable ( $X = \{x^i\}_1^N$ , where  $N$  is the number of samples). The framework of this strategy is shown in Figure 1, where  $x_i^{(j)}$  represents the  $j$ th feature of the  $i$ th sample.

First, a data point  $x_i$  is input into the encoder. Through the neural network, we obtain the parameters of the approximate posterior distribution  $q_\phi(z|x_i)$  obeyed by the hidden variable  $z$ .

The posterior distribution is a Gaussian distribution, and the output of the encoder includes the parameters  $\sigma_i^2$  and  $\mu_i$  of the Gaussian distribution that  $z|x_i$  obeys. With the parameters  $\sigma_i^2$  and  $\mu_i$  of the  $z|x_i$  distribution, we sample one  $\epsilon_i$  from  $\mathcal{N}(0, I)$  and use the reparameterization trick to set  $z_i = \mu_i + \sigma_i \square \epsilon_i$  where  $z_i$  represents the value of a similar sample  $x_i$ ; and  $\square$  represents the elementwise multiplication operation. The decoder needs to fit the likelihood distribution  $p_\theta(X|z_i)$  and feed a  $z_i$  to the decoder, which returns the parameters of the distribution that  $X|z_i$  obeys; the likelihood also obeys a Gaussian distribution. After obtaining the parameters of the distribution of  $|z_i$ , we sample from the distribution to generate a sample  $x_i$ . In the last step, we do not sample and directly regard the  $\mu'_i$  output by the model as the sample  $x_i$  generated by  $z_i$ . Then, we can obtain the objective function of the VAE, and we only need to maximize  $\mathcal{L}$ .

$$\mathcal{L}(p_\theta, q_\phi) = -D_{KL}(q_\phi, p) + \mathbb{E}_{q_\phi}[\log p_\theta(X|z)]$$

$-D_{KL}(q_\phi, p)$  is the Kullback-Leibler (KL) divergence of the two distributions  $p$  and  $q$  and is also a regular term.  $\mathbb{E}_{q_\phi}[\log p_\theta(X|z)]$  is often referred to as the reconstruction loss.



**FIGURE 2**  
The graphical structure of the VAE model for DPLs prediction.

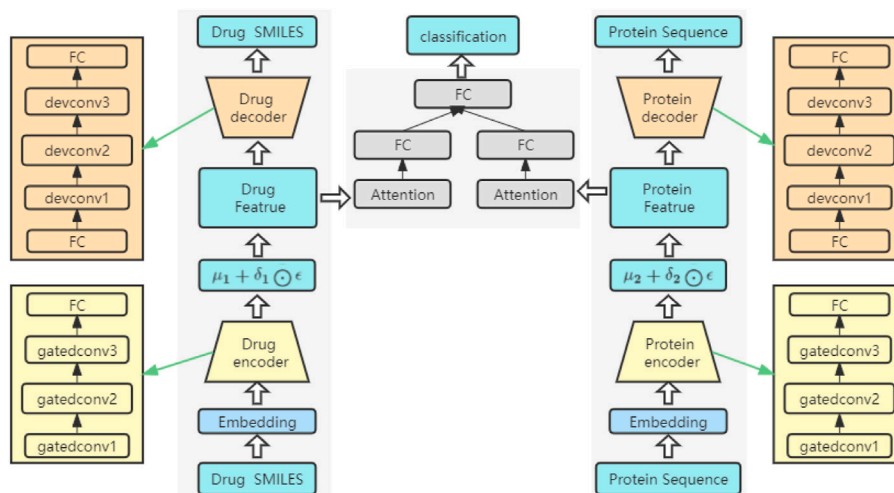


FIGURE 3  
The framework structure of the model.

With the theoretical support of the VAE, we apply it to the prediction of DPIs. Here,  $x_d$  and  $x_t$  represent the drug and protein vectors, respectively, and  $y$  represents the interaction relationship or affinity value between drug and protein. We assume that the hidden variables of  $x_d$  and  $x_t$  are  $z_d$  and  $z_t$ , respectively. Our aim is to learn a model that predicts drug and protein interactions.

The model diagram for applying variational autoencoding to DPIs prediction is shown in Figure 2. The model has two encoders, which are mainly used to generate latent variables  $z_d$  and  $z_t$  for drug  $x_d$  and protein  $x_t$ , respectively. Three important decoders are used to generate  $x_d$  and  $x_t$  from the latent variables  $z_d$  and  $z_t$  and generate the drug-protein relationship  $y$ . Finally, for each drug-protein pair, the objective function of the VAE model is to maximize  $\mathcal{L}'$ .

$$\begin{aligned} \mathcal{L}' &= \mathcal{L}_{DrugVAE} + \mathcal{L}_{ProteinVAE} \\ &= \left\{ -D_{KL}(q_{\varnothing_d}, p_d) + \mathbb{E}_{q_{\varnothing_d}}[\log p_{\theta_d}(X_d|z_d)] \right\} \\ &\quad + \left\{ -D_{KL}(q_{\varnothing_t}, p_t) + \mathbb{E}_{q_{\varnothing_t}}[\log p_{\theta_t}(X_t|z_t)] \right\} \end{aligned}$$

## Attention mechanism for feature extraction

Attention mechanisms, as effective means of feature screening and enhancement, have been widely used in many fields of deep learning. A structural model based on an attention mechanism can not only record the positional relationships between pieces of information but also measure the importance levels of different information features according to the weight of the information. Dynamic

weight parameters are established by making relevant and irrelevant choices for the information features to strengthen the key information and weaken the useless information, thereby improving the efficiency of deep learning algorithms and improving some of the defects of traditional deep learning techniques.

Utilizing an attention mechanism for the prediction of DPIs can enable effective atomic feature extraction because the structures of the molecular sequences of drugs and proteins are very similar to the structures of natural language sentences, and the context information of atoms is very important for understanding molecular features (Jastrzębski et al., 2018). In detail, we should pay attention to the interaction information of each atom and its adjacent atoms; each atom is also connected to the simplified molecular-input line-entry system (SMILES (Weininger, 1988), which is a symbol for molecular structure encoding). Information about the interactions of atoms that are farther away in the sequence can also have an impact on the predicted results. The molecular sequences of proteins are very long, and the best way to extract features is to use an attention mechanism.

Attention mechanisms are widely used in the natural language processing (NLP) field, and they have also been shown to be powerful for processing textual data. The core of such a mechanism is an attention function (Vaswani et al., 2017). The attention function can be described as mapping a query (Q) and a set of key-value (K-V) pairs to an output. Among them, dot product attention with Q, K and V is a widely used attention approach. Let the dimensions of Q and K be  $d_k$  and the dimensionality of V be  $d_v$ . Then, attention can be expressed by the following formula.

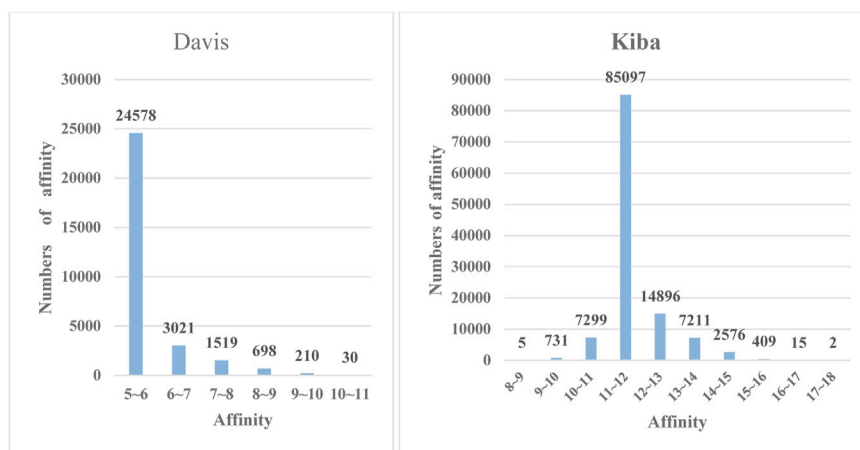


FIGURE 4

The frequency histograms of the affinities in the Davis and KIBA datasets. The horizontal axis denotes the affinity values of drugs and proteins, and the vertical axis represents the numbers of affinity values in certain intervals.

TABLE 1 Model parameters.

Parameter	Value
Number of filters in the encoder and decoder	32
Filter length (drug molecules)	5
Filter length (protein sequences)	7
Number of epochs	300
Batch size	256
Learning rate	0.001, 0.0001

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where  $Q \in \mathbb{R}^{n \times d_k}$ ,  $K \in \mathbb{R}^{m \times d_k}$  and  $V \in \mathbb{R}^{m \times d_v}$ . The attention measures the similarity between the inner product of the matrix  $Q$  and  $V$ , which is nonlinearly transformed by a softmax function with the matrix  $V$ . Here,  $d_k$  prevents the forward propagation of the network due to excessive data content.

## The network structure of the model

The network structure of the model is shown in Figure 3. It consists of three key parts: an encoder, a decoder and a prediction module. Both the encoder and the decoder serve to predict the interactions between drugs and proteins. Among them, the overall structure extracts the features of drugs and proteins, sends the extracted features into the attention block to focus on the important parts, and finally sends them to the FC layer to predict the DPIs. The feature extraction process for drugs is the

same as that for proteins. Before being fed into the encoder, drugs and proteins are sequences of text strings, which need to be converted into digital vectors. According to the existing character dictionary, each character is converted into an integer type, and then each sample is converted into an embedding matrix through embedding. In our model, three GatedCNNs are included in the coding layer, and a rectified linear unit (RELU) (Nair and Hinton, 2010) activation function is present after each layer of CNNs. The filters of the last two CNNs are the first CNNs, which are filtered two times and three times. A max pooling layer is appended after the third GatedCNN to compress the extracted features.

In the decoders for drugs and proteins, the input source data are reconstructed through deconvolutional networks (Zeiler et al., 2010). Each decoder has an FC layer and three deconvolution layers. The last deconvolution appends an FC layer to convert the output into drug and target sequences with the same size as that of the input.

In the DPIs prediction module, two FC layers are used to represent the features of drugs and proteins. To further extract the high-level features of drugs and proteins, a self-attention mechanism is introduced after the FC layers, focusing on important features in drug sequences or protein sequences and ignoring unnecessary features. Then, the final extracted high-level features are spliced and sent to a network containing three FC layers. A ReLU activation function and a dropout function are placed after the first two FC layers, and the dropout function is mainly used to prevent the network from overfitting. The final output can be used to predict DPIs.

The model parameters used in this experiment are shown in Table 1. Among them, we select several values [16,32,64] for the number of CNN filters in the encoder and decoder and find that

TABLE 2 AUC, precision, recall, and F1 values obtained under different methods.

Models	AUC	Precision	Recall	F1
BindingDB dataset				
K-NN	0.776	0.762	0.791	0.776
RF	0.742	0.834	0.600	0.698
L2	0.737	0.784	0.646	0.709
SVM	0.805	0.770	<u>0.858</u>	0.811
BridgeDPI	<b>0.960</b>	<u>0.883</u>	<b>0.903</b>	<b>0.893</b>
Ours	<u>0.913</u>	<b>0.888</b>	0.822	<u>0.854</u>
C.elegans dataset				
K-NN	0.858	0.801	0.827	0.814
RF	0.902	0.821	0.844	0.832
L2	0.892	0.890	0.877	0.883
SVM	0.894	0.785	0.818	0.801
BridgeDPI	<b>0.996</b>	<b>0.980</b>	<b>0.970</b>	<b>0.975</b>
Ours	<u>0.925</u>	<u>0.927</u>	<u>0.897</u>	<u>0.912</u>
Human dataset				
K-NN	0.860	0.798	0.927	0.858
RF	<u>0.940</u>	0.861	0.897	0.879
L2	0.911	0.891	0.913	0.902
SVM	0.910	<b>0.966</b>	<u>0.950</u>	<u>0.958</u>
BridgeDPI	<b>0.990</b>	<u>0.962</u>	<b>0.965</b>	<b>0.963</b>
Ours	0.914	0.934	0.862	0.897

The best result was highlighted in bold, while the suboptimal is denoted by underline.

the effect of 32 filters was best and that the filter lengths of drugs and proteins are both in [5,7,9,11]. We choose the best results, and the final filter lengths of drugs and proteins are 5 and 7, respectively.

## Experiment

### Datasets

To verify the effectiveness of the proposed model and compare it with the base method, we conducted experiments on the following datasets: C.elegans and Human datasets, BindingDB dataset, Davis dataset and KIBA dataset.

#### C.elegans and human datasets

In the work of (Liu et al., 2015), the authors used a systematic scanning framework, and their dataset contained a large number of negative samples. They constructed two datasets, C.elegans and human. Following the requirements of (Tsubaki et al., 2019), we used a balanced dataset with an approximately 1:1 ratio of positive and negative samples. The C.elegans dataset includes 1876 protein targets and 1767 drug molecules, and it contains 7786 affinity sample pairs, 3893 positive samples, and 3893 negative samples. The human dataset contains

6728 affinity pairs, the number of protein targets is 2001, and the number of drug molecules is 2726.

#### BindingDB dataset

BindingDB is a public, web-accessible database of measured binding affinities that focuses chiefly on the interactions of proteins considered to be drug targets with small, drug-like molecules. In this experiment, the method described in the paper of (Gao et al., 2018) was used; the dataset contains 39,747 positive samples and 31,218 negative samples.

#### Davis

The Davis dataset contains affinity pairs measured by their  $K_d$  value (kinase dissociation constant); it includes 68 drug molecules and 442 proteins (Davis et al., 2011). The  $K_d$  value can reflect the affinity between a drug and a protein. It is a bridge for predicting affinity, and its affinity value range is [0.016, 10000]. The frequency distribution plot of affinity values for the Davis dataset is shown in Figure 4.

#### KIBA

Measuring the relationships between the drugs and proteins in the KIBA dataset is mainly achieved through the KIBA score (Tang et al., 2014) combined the biological activities of different sources of kinase inhibitors, such as  $K_d$ ,  $K_i$  and  $IC_{50}$ , to calculate KIBA affinity scores. We conducted experiments according to this environment. Then, we found that the dataset contains 52498 drug molecules, 476 protein sequences, and 246088 KIBA scores and that the numerical range of the KIBA score is [0, 17.2]. Values in this dataset with KIBA scores that were less than 10 were removed, leaving 2111 drug molecules and 229 proteins in the end. The frequency distribution plot of the affinity values for the KIBA dataset is shown in Figure 4.

### Training details

The model was implemented based on Python 3.6 and PyTorch 1.10.2. The program ran on a GTX1060 GPU with 8 GB of memory. The network parameter initialization process was implemented by the *xavier\_normal\_()* function in the library. During training, the network used the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.0001 for the Davis and KIBA datasets and a learning rate of 0.001 for the other datasets to adjust the network parameters. To prevent overfitting, L2 regularization was added to the loss function. Each batch contained 256 samples, and the samples were randomly scrambled. Three hundred epochs were executed. Finally, the model was trained by minimizing the cross-entropy loss function.

$$L(y, \hat{y}) = -(y \log \hat{y} + (1 - y) \log (1 - \hat{y})) + \lambda \sum \|\Theta\|^2$$



where  $y$  is the true value,  $\hat{y}$  is the predicted value,  $\lambda$  is the regularization coefficient, and  $\Theta$  is the network parameter. On the basis of the parameters in Table 1, we search the optimal value of  $\lambda$  by grid searching scheme within range of  $[-5, +5]$ . Empirical experiments show that setting the value of  $\lambda$  being  $-3$  yield the best performances.

## Results

First, we conducted experiments on the BindingDB dataset extracted by Gao et al. According to the environment they set for the dataset, we utilized the same division to ensure that the data in the validation set did not appear in the training set, so that the experiment was closer to the real-world situation. During the training process, to prevent overfitting, we set the termination criterion according to the ACC evaluation index of the validation set. When the ACC of the validation set iterated for a certain number of steps and did not increase, the program terminated. To demonstrate the superiority of the model, we made a comparison with k-nearest neighbors (K-NN), a random forest (RF), L2, a support vector machine (SVM) and the BridgeDPI model (Wu et al., 2022a). The machine learning results for these methods were derived from the source paper on C.elegans and Human dataset (Tsubaki et al., 2019). We conducted experiments on BindingDB dataset. The table shows that on the BindingDB dataset, the AUC, Precision, Recall, and F1 of the proposed model reached 0.913, 0.888, 0.822, and 0.854, respectively. Our model outperforms traditional machine learning methods in AUC, Precision, and F1. The unsupervised K-NN method yielded lower results than the other models and methods, with AUCs and F1 scores of 0.858/0.814 and 0.860/0.858 on the C.elegans and human datasets, respectively. The effects of the RF, L2, and the SVM based on supervised learning were better. The AUC on the C.elegans dataset reached approximately 0.9, and the AUC on the human data exceeded 0.9. Compared with traditional machine learning methods, our model achieved the highest evaluation indicators on the C.elegans dataset, and its AUC, precision, recall, and F1 were 2.3%, 3.7%, 2.0%, and 2.9% higher than those of the second-best approaches, respectively. At the same time, our method performed slightly better on the human dataset. Since models such as K-NN, the RF, L2, and the SVM cannot obtain high-quality feature information, it is not easy for them to learn complex nonlinear DPIs. However, deep learning has strong feature extraction capabilities. Our model benefits from that. The Table 2 shows that our model doesn't perform as well as BridgeDPI that extracts features from the biological perspective. Our model is similar to nature language processing in extracting features, and it's indeed not as effective as BridgeDPI. However, our model has some advantages: when dealing with drug features and protein features, an attention mechanism is introduced to realize the key sites of drug-protein binding, thereby ignoring irrelevant site information and saving the screening time of drug-

TABLE 3 Multi-classification results obtained on the Davis and KIBA datasets.

Models	ACC		AUC	
	Davis	KIBA	Davis	KIBA
K-NN	0.854	0.777	0.581	0.574
RF	<b>0.868</b>	<u>0.811</u>	<u>0.610</u>	<u>0.591</u>
L2	0.752	0.699	0.582	0.556
SVM	<u>0.862</u>	0.801	0.541	0.543
Ours	0.850	<b>0.841</b>	<b>0.705</b>	<b>0.813</b>

The best result was highlighted in bold, while the suboptimal is denoted by underline.

protein interactions. This has contributed to experts to identify drug-protein interactions.

To further demonstrate the feature extraction advantages of deep models, we performed a multiclass prediction experiment on the Davis and KIBA datasets, and the results are shown in Table 3. The Davis and KIBA datasets possess continuous values, and the  $pK_d$  values of Davis and the scores of KIBA are distributed as shown in Figure 4. We found their means  $\mu$  and variances  $\sigma$  according to the relationship between the mean and variance. We divided the data into five categories:  $[\mu - \sigma, \mu + \sigma]$ ,  $[\mu - 2\sigma, \mu + 2\sigma]$ ,  $[\mu - 3\sigma, \mu + 3\sigma]$ ,  $[\mu - 4\sigma, \mu + 4\sigma]$ , and *other*. According to our model, the classification effect of the test was relatively objective, and the ACC and AUC reached 0.850/0.705 and 0.841/0.813 on the Davis and KIBA datasets, respectively. AUC is the best on both Davis and KIBA datasets. ACC is lower on Davis dataset due to uneven data distribution and less data, which may affect our results. The results show that the proposed model is robust.

## Conclusion

In this work, we propose a model based on VAEs and attention mechanisms to predict DPIs. The high-level features of drugs and proteins are further extracted by a CNN and an attention mechanism. Experiments show that our method outperforms some base methods on the testing datasets and illustrates the powerful ability of deep learning to extract features. To further verify the robustness of the model, we perform a multiclass prediction experiment on the Davis and KIBA datasets. The final results of the experiment yield good metric values.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.bindingdb.org/bind/index.jsp>, <http://staff.cs.utu.fi/~aatapa/data/DrugTarget/>, [https://wormbase.org/species/c\\_elegans#104-10](https://wormbase.org/species/c_elegans#104-10).

## Author contributions

YZ: Ideas; Formulation of overarching research goals and aims; Development or design of methodology; Creation of models; Data analysis; Review and revision of the first draft. YH: Implementation of the computer code; Original Draft; Preparation of experimental data. HL: Formal analysis; Preparation of experimental data; Review and revision of the first draft. XL: Experimental Supervision and Leadership; Review and revision of the first draft.

## Funding

This work was partly supported by National Natural Science Foundation of China (62172112, 62172113).

## References

- Avorn, J. (2015). The \$2.6 billion pill — methodologic and policy considerations. *N. Engl. J. Med.* 372, 1877–1879. doi:10.1056/NEJMp1500848
- Ballester, P. J., and Mitchell, J. B. O. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 1169–1175. doi:10.1093/bioinformatics/btq112
- Bleakley, K., and Yamanishi, Y. (2009). Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* 25, 2397–2403. doi:10.1093/bioinformatics/btp433
- Chen, J., Peng, H., Han, G., Cai, H., and Cai, J. (2019). Hogmmnc: A higher order graph matching with multiple network constraints model for gene–drug regulatory modules identification. *Bioinformatics* 35, 602–610. doi:10.1093/bioinformatics/bty662
- Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., et al. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 29, 1046–1051. doi:10.1038/nbt.1990
- Faulon, J.-L., Misra, M., Martin, S., Sale, K., and Sapra, R. (2008). Genome scale enzyme–metabolite and drug–target interaction predictions using the signature molecular descriptor. *Bioinformatics* 24, 225–233. doi:10.1093/bioinformatics/btm580
- Gao, K. Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S., and Zhang, P. (2018). “Interpretable drug target prediction using deep neural representation,” in *Proceedings of the twenty-seventh international joint conference on artificial intelligence, twenty-seventh international joint conference on artificial intelligence [IJCAI-18]* (Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization), 3371–3377.
- Gschwend, D. A., Good, A. C., and Kuntz, I. D. (1996). Molecular docking towards drug discovery. *J. Mol. Recognit.* 9, 175–186. doi:10.1002/(sici)1099-1352(199603)9:2<175::aid-jmr260>3.0.co;2-d
- Huang, J., Chen, J., Zhang, B., Zhu, L., and Cai, H. (2021). Evaluation of gene–drug common module identification methods using pharmacogenomics data. *Brief. Bioinform.* 22, bbaa087. doi:10.1093/bib/bbaa087
- Jastrzębski, S., Leśniak, D., and Czarnecki, W. M. (2018). *Learning to SMILE(S)*. Tokyo, Japan: The Institute of Electronics, Information and Communication Engineers.
- Keum, J., and Nam, H. (2017). SELF-BLM: Prediction of drug–target interactions via self-training SVM. *PLOS ONE*. *PLoS One* 12, e0171839. doi:10.1371/journal.pone.0171839
- Kingma, D., and Ba, J. (2014). *Adam: A method for stochastic optimization* in 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings. International Conference on Learning Representations, ICLR (San Diego, CA: OpenReview.net).
- Kingma, D. P., and Welling, M. (2014). *Auto-encoding variational bayes*. in 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings. International Conference on Learning Representations, ICLR (Banff, Canada: OpenReview.net).
- Liu, H., Sun, J., Guan, J., Zheng, J., and Zhou, S. (2015). Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics* 31, i221–i229. doi:10.1093/bioinformatics/btv256
- Liu, M.-Y., Breuel, T., and Kautz, J. (2018). *Unsupervised image-to-image translation networks* in Advances in neural information processing systems (Red Hook, NY, United States: Curran Associates Inc), 700–708.
- Liu, Y., Wu, M., Miao, C., Zhao, P., and Li, X.-L. (2016). Neighborhood regularized logistic matrix factorization for drug–target interaction prediction. *PLOS Comput. Biol.* 12, e1004760. doi:10.1093/bioinformatics/btaa577
- Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M., Wegner, K. J., Ceulemans, H., et al. (2018). Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* 9, 5441–5451. Royal Society of Chemistry. doi:10.1039/c8sc00148k
- Min, S., Lee, B., and Yoon, S. (2016). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869. doi:10.1093/bib/bbw068
- Nascimento, A. C. A., Prudêncio, R. B. C., and Costa, I. G. (2016). A multiple kernel learning algorithm for drug–target interaction prediction. *BMC Bioinforma.* 17, 46. doi:10.1186/s12859-016-0890-3
- Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., and Venkatesh, S. (2019). *GraphDTA: Predicting drug–target binding affinity with graph neural networks*, preprint. *Bioinformatics* 37, 1140–1147. doi:10.1093/bioinformatics/btaa921
- Oprea, T. I., and Mestres, J. (2012). Drug repurposing: Far beyond new targets for old drugs. *AAPS J.* 14, 759–763. doi:10.1208/s12248-012-9390-1
- Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics* 34, i821–i829. doi:10.1093/bioinformatics/bty593
- Öztürk, H., Ozkirimli, E., and Özgür, A. (2019). *WideDTA: Prediction of drug–target binding affinity*. (Tokyo, Japan: The Institute of Electronics, Information and Communication Engineers).
- Qiu, Y., Ching, W.-K., and Zou, Q. (2021a). Matrix factorization-based data fusion for the prediction of RNA-binding proteins and alternative splicing event associations during epithelial–mesenchymal transition. *Brief. Bioinform.* 22, bbab332. doi:10.1093/bib/bbab332
- Qiu, Y., Ching, W.-K., and Zou, Q. (2021b). Prediction of RNA-binding protein and alternative splicing event associations during epithelial–mesenchymal transition based on inductive matrix completion. *Brief. Bioinform.* 22, bbaa440. doi:10.1093/bib/bbaa440
- Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K., et al. (2014). Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *J. Chem. Inf. Model.* 54, 735–743. doi:10.1021/ci400709d
- Tian, K., Shao, M., Wang, Y., Guan, J., and Zhou, S. (2016). Boosting compound–protein interaction prediction by deep learning. *Methods* 110, 64–72. doi:10.1016/j.ymeth.2016.06.024

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



- Tsubaki, M., Tomii, K., and Sese, J. (2019). Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences', *Bioinformatics*. *Bioinformatics* 35, 309–318. doi:10.1093/bioinformatics/bty535
- Unterthiner, T., and Mayr, A. (2014). *Deep learning as an opportunity in virtual screening* in Proceedings of the deep learning workshop at NIPS (Cambridge, MA, United States: MIT Press), 27, 1–9.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). *Attention is all you need* in Advances in Neural Information Processing Systems30 (NIPS) (Red Hook, NY, United States: Curran Associates Inc), 5998–6008.
- Walker, J., Marino, K., Gupta, A., and Hebert, M. (2017). "The pose knows: Video forecasting by generating pose futures," in *2017 IEEE international conference on computer vision (ICCV)* (Venice: IEEE), 3352–3361.
- Wang, L., You, Z.-H., Chen, X., Xia, S.-X., Liu, F., Yan, X., et al. (2018). A computational-based method for predicting drug–target interactions by using stacked autoencoder deep neural network. *J. Comput. Biol.* 25, 361–373. doi:10.1089/cmb.2017.0135
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* 28, 31–36. doi:10.1021/ci00057a005
- Wu, Y., Gao, M., Zeng, M., Zhang, J., and Li, M. (2022a). BridgeDPI: A novel graph neural network for predicting drug–protein interactions', *bioinformatics*. *Bioinformatics* 38, 2571–2578. doi:10.1093/bioinformatics/btac155
- Wu, Y., Zeng, M., Fei, Z., Yu, Y., Wu, F.-X., and Li, M. (2022b). Kaicd: A knowledge attention-based deep learning framework for automatic icd coding. *Neurocomputing* 469, 376–383. doi:10.1016/j.neucom.2020.05.115
- Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S. (2010). Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26, i246–i254. doi:10.1093/bioinformatics/btq176
- Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R. (2010). "Deconvolutional networks," in *2010 IEEE computer society conference on computer vision and pattern recognition, 2010 IEEE conference on computer vision and pattern recognition (CVPR)* (San Francisco, CA, USA: IEEE), 2528–2535.
- Zeng, M., Zhang, F., Wu, F.-X., Li, Y., Wang, J., and Li, M. (2019). Protein–protein interaction site prediction through combining local and global features with deep neural networks. *Bioinformatics* 36, 1114–1120. doi:10.1093/bioinformatics/btz699
- Zhang, F., Song, H., Zeng, M., Li, Y., Kurgan, L., and Li, M. (2019). DeepFunc: A deep learning framework for accurate prediction of protein functions from protein sequences and interactions. *PROTEOMICS* 19, 1900019. doi:10.1002/pmic.201900019
- Zhang, F., Song, H., Zeng, M., Wu, F.-X., Li, Y., Pan, Y., et al. (2021). A deep learning framework for gene ontology annotations with sequence- and network-based information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 2208–2217. doi:10.1109/TCBB.2020.2968882