



Methods

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

DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions

Tao Song^{a,c,*}, Xudong Zhang^a, Mao Ding^{b,*}, Alfonso Rodriguez-Paton^c, Shudong Wang^a, Gan Wang^a

^a College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China

^b Department of Neurology Medicine, The Second Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250033, China

^c Department of Artificial Intelligence, Faculty of Computer Science, Polytechnical University of Madrid, Campus de Montegancedo, Boadilla del Monte 28660, Madrid, Spain

ARTICLE INFO

Keywords:

Drug-target interaction
Feature extraction
Multi-scale fusion
Deep learning

ABSTRACT

Predicting drug-target interactions (DTIs) is essential for both drug discovery and drug repositioning. Recently, deep learning methods have achieved relatively significant performance in predicting DTIs. Generally, it needs a large amount of approved data of DTIs to train the model, which is actually tedious to obtain. In this work, we propose DeepFusion, a deep learning based multi-scale feature fusion method for predicting DTIs. To be specific, we generate global structural similarity feature based on similarity theory, convolutional neural network and generate local chemical sub-structure semantic feature using transformer network respectively for both drug and protein. Data experiments are conducted on four sub-datasets of BIOSNAP, which are 100%, 70%, 50% and 30% of BIOSNAP dataset. Particularly, using 70% sub-dataset, DeepFusion achieves ROC-AUC and PR-AUC by 0.877 and 0.888, which is close to the performance of some baseline methods trained by the whole dataset. In case study, DeepFusion achieves promising prediction results on predicting potential DTIs in case study.

1. Introduction

Prediction of drug-target protein interactions (DTIs) is critical in the field of drug discovery and drug repositioning [1]. It has relied heavily on hundreds of currently known drug targets to detect drugs [2]. Although existing bioassay technologies can be used for DTIs detection, large-scale experiments still have many limitations. In addition, the cost of the experiment and scarcity of public drug repositioning analysis data made it necessary to develop suitable computational tools to accurately detecting DTIs [3]. Plenty of methods have been developed for predicting DTIs, such as ligand-based methods, docking methods and feature learning methods. Most ligand-based methods are lying on the basis of quantitative structure–activity relationship (QSAR), which assumes molecules with similar structure have similar biological activities [4]. The highlight of ligand-based methods is characterized by molecular similarity of drug-target to predict interactions, which is inaccurate and not generalizable. These problems can be avoided to some extent by building multi-target and fragment-based QSAR models [5–7]. Docking is one of the most commonly used methods in the field of drug-target

interaction prediction, which models in crystal structure of target to screen small molecules [8]. The DTIs predicted by docking methods are relatively accurate, however most methods to obtain 3D structural information on target such as X-ray crystallography are rather time-consuming and expensive [9].

Feature learning methods can map drug and target into numerical descriptors, combined with machine learning or deep learning methods such as random forest (RF) [10], support vector machine (SVM) [11] and convolutional neural network (CNN) [12–14] to learn potential relationship of drug-target pairs, which regards the prediction of DTIs as a binary classification problem. Recently, a number of deep learning models [15–17] for prediction of DTIs have emerged.

All of them have improved accuracy to a greater or lesser extent, but there are still several common unresolved problems as follows:

- **Scarcity of known drug-target interaction leading to limited application scenarios for models.** Recently, many transformer and graph-based methods have emerged. Graph-DTA [18] uses multiple GNNs to capture graph information of drug and CNN to capture

* Corresponding authors at: College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China (T. Song).

E-mail addresses: tsong@upc.edu.cn (T. Song), 18264181312@163.com (M. Ding).

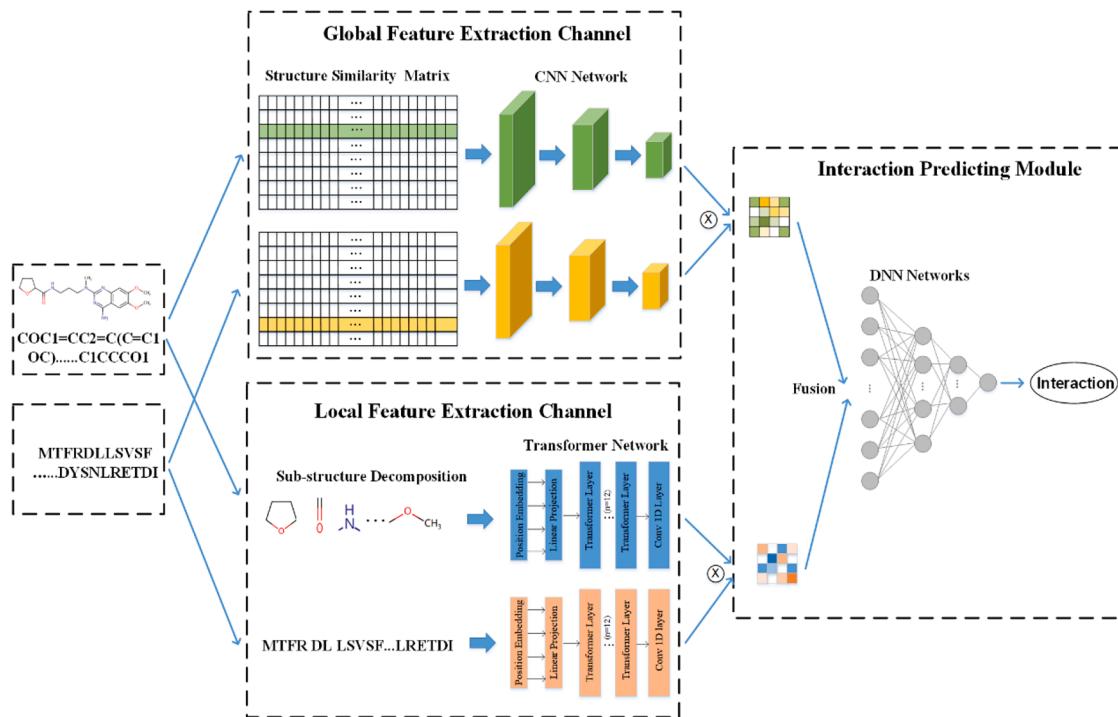


Fig. 1. The Overview of DeepFusion.

sequence information of protein. Transformer-CPI [19] performs DTI prediction by feature extraction of drug and protein sequences. All of them require large amounts of DTIs data to train the model. However, the majority of DTIs data is unlabeled and only a small proportion of known DTIs data is available. In addition, when we try to implement some DTI models on specific cases, such as diabetes, the size of available training data is even smaller.

- **Limited to encoding molecular feature and ignoring how to model interaction feature.** Most current deep learning methods [20,21] only focus on encoding molecule feature through extracted molecular feature to predict interactions, which in fact extracts redundant feature that interfere and disrupt prediction performance. Moltrans [22] extracts drug and protein sub-structural feature to avoid the aforementioned problems like redundant feature and etc. However, the dimensionality of final drug-protein interaction feature is too large and only considered from single perspective, thus it is not comprehensive enough to accurately predict small samples of DTIs data.

In this work, we tackled two challenging problems by proposing a deep learning based multi-scale interaction feature fusion method for predicting DTIs. The main contributions of this paper are as follows: (1) In previous prediction of DTIs, the structure of a drug or protein is embedded using only its own information, without considering the global feature with other molecules. And this paper proposes a method to obtain more information in limited data to make accurate predictions. Tanimoto coefficient, Levenshtein distance and convolutional neural network are applied to generate drug and protein global structural similarity features. (2) We not only consider the global feature of molecule, but also note that drug-target interaction occurs on molecular sub-structures. In our work, we use transformer to extract the sub-structure feature of molecules. (3) Two feature extraction channels are used to obtain global feature based interaction and local sub-structure feature based interaction, the two features are fused for the final prediction.

2. Materials and methods

2.1. Problem definition

We take DTIs prediction problem here as a classification task by using Simplified Molecular Input Line Entry System (SMILES) for drug and protein sequence as input to predict DTIs. Drug and protein datasets are denoted by Φ^D and Φ^P respectively. Inputs to our model are individual drug and protein denoted by Φ_i^D and Φ_j^P respectively as well. In order to get the structural similarity feature of Φ_i^D , all remaining drugs in Φ^D are the benchmark set for current drug. And the same is true for proteins. The benchmark set is represented by Ω^D and Ω^P for drug and protein respectively, where individual drug and protein in the benchmark datasets are represented by Ω_i^D and Ω_j^P .

2.2. The DeepFusion method

DeepFusion consists of global structural similarity feature extraction channel (Sim-Channel), locals ub-structure feature extraction channel (Sub-Channel) and interaction predicting module (IP-Module). The Overview of the topological structure of DeepFusion is shown in Fig. 1. Sim-Channel operates on pre-generated drug/protein structural similarity matrix to obtain drug/protein structural similarity feature (DSSF/ PSSF) by using similarity theory and convolutional neural network. Sub-Channel uses Frequent Consecutive Sub-sequence (FCS) [22] mining that is adaptable to extract high-quality fifite-sized sub-structure for both protein and drug. The contextual embedding of sub-structure was subsequently enhanced by using transformer [23] to learn sub-structure information and chemical semantic information of input sequence. Dimensionality reduction is performed using CNNs in the last layer of transformer network to avoid dimensionality explosion. Interaction predicting module (IP-Module) fuses these two parts of features, generating a final encoded interaction feature for more accurate prediction to obtain the probability score of predicted DTI.

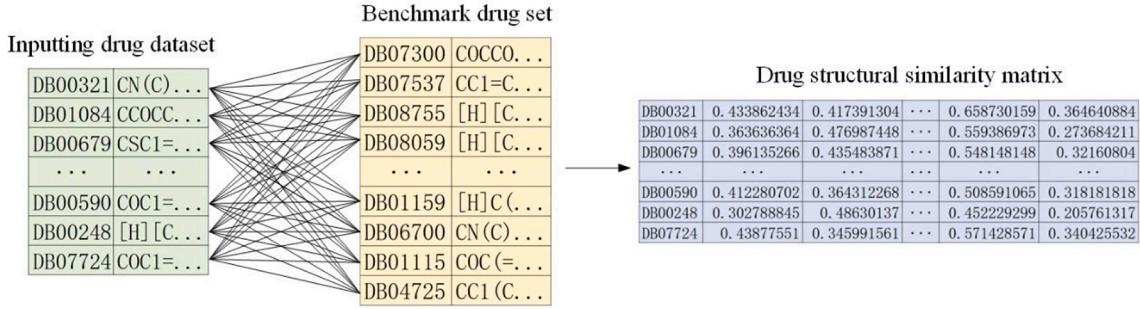


Fig. 2. Generation of drug structure similarity matrix.

2.2.1. Extraction of global structural similarity feature

In order to structurally relate each drug to all other drugs in same dataset, we need to calculate structural similarity of each drug based on structural information of all other drugs, which will generate a drug structural similarity matrix. Similarly, for proteins, a protein structure similarity matrix is generated. Sim-Channel is first implemented with the generated drug and protein structural similarity matrices. The individual drug and protein input to Sim-Channel are then matched on drug and protein structural similarity matrices respectively. DSSF and PSSF of input drug and protein are then obtained. For DSSF, it contains calculated structural similarity scores of input drug to all drugs in benchmark dataset. DSSF can greatly capture structural associations of individual drug with all drugs in benchmark dataset, enabling more accurate representation of input individual drug.

- Generation of drug structural similarity matrix (matrix M^D)

DeepFusion uses SMILES of drug as input, while Sim-Channel first uses Rdkit to convert drug SMILES to Morgan Fingerprint with radius of 2 in preparation for the subsequent calculation of drug structural similarity. Morgan Fingerprint represent drug as a one-hot vector, indicating the presence or absence of different sub-structures or pharmacological features at each location. We use Tanimoto coefficient (TC) to quantify similarity extent between two Morgan Fingerprint. The similarity between Morgan Fingerprint A and Morgan Fingerprint B is calculated as follows [24]:

$$\text{Sim}(MF_A, MF_B) = \frac{|MF_A \cap MF_B|}{|MF_A \cup MF_B|} \quad (1)$$

where $|MF_A \cap MF_B|$ is the number of features that Morgan Fingerprint A and Morgan Fingerprint B have in common and $|MF_A \cup MF_B|$ is the number of all features of Morgan fingerprint A and Morgan fingerprint B. $\text{Sim}(MF_A, MF_B)$ takes on a value in the range 0 to 1. When $\text{Sim}(MF_A, MF_B) = 1$, it means that two drugs are perfectly similar. When $\text{Sim}(MF_A, MF_B) = 0$, it means that the two drugs are approximately completely unrelated.

Two-by-two structural similarity calculation are cyclically performed for drug Φ_i^D in dataset Φ^D and drug Ω_j^D in benchmark set Ω_j^D to generate a drug structural similarity matrix M^D .

$$M^D = \begin{pmatrix} \text{Sim}(MF_{\Phi_1^D}, MF_{\Omega_1^D}) & \dots & \text{Sim}(MF_{\Phi_1^D}, MF_{\Omega_j^D}) \\ \vdots & \ddots & \vdots \\ \text{Sim}(MF_{\Phi_i^D}, MF_{\Omega_1^D}) & \dots & \text{Sim}(MF_{\Phi_i^D}, MF_{\Omega_j^D}) \end{pmatrix} \quad (2)$$

where the i-th row in M^D represents structural similarity feature vector V_i^{DSSF} of the i-th drug Φ_i^D in input dataset. $\text{Sim}(MF_{\Phi_i^D}, MF_{\Omega_j^D})$ is similarity score of the i-th drug in input data set to the j-th drug in benchmark dataset. Drug structural similarity matrix M^D establishes a structural link between input individual drug and all drugs in benchmark set.

- Generation of protein structural similarity matrix (matrix M^P)

We use Levenshtein distance [25] to calculate protein sequences structural similarity. Levenshtein distance between two protein sequences calculates the minimum number of substitutions required to convert one sequence to another by means of insertion, deletion and replacement character operations. The similarity of two protein sequences is measured by the size of the Levenshtein distance, with larger values indicating that two protein sequences are more similar, and conversely smaller values indicating that two protein sequences are less similar. Levenshtein distance allows for comparison of sequences of different lengths, so it can be used well for similarity comparisons of structurally and functionally different protein sequences.

Suppose there is two protein sequences:

$P = \{\text{MVKFNNSSRKSGSKKTIRK...RIVYIGEHKNQKV}\}$ of length $|P|$,

$Q = \{\text{MTELKAKGPGRAPHVAGGPP...LPKILAGMVVKPLLFHKK}\}$ of length $|Q|$.

To calculate Levenshtein distance of P and Q, first create a matrix L $\in \mathbb{R}^{(|P|+1) \times (|Q|+1)}$ and then perform the following calculation:

$$L_{P,Q}(i,j) = \begin{cases} \max(j,j) & \\ \min \begin{cases} L_{P,Q}(i-1,j) + 1 & \text{if } \min(i,j) = 0 \\ L_{P,Q}(i,j-1) + 1 & \text{otherwise} \\ L_{P,Q}(i-1,j-1) + 1 & (P_i \neq Q_j) \end{cases} & \end{cases} \quad (3)$$

For $L_{P,Q}(i,j)$, if $i = 0$ and $j = 0$, then $L_{P,Q}[0,0] = 0$. If $i = 0$, then $L_{P,Q}[0,j] = j$. If $j = 0$, then $L_{P,Q}[i,0] = j$. If $i \geq 1$ and $j \geq 1$, then $L_{P,Q}[i,j] = \min\{L_{P,Q}[i-1,j] + 1, L_{P,Q}[i,j-1] + 1, L_{P,Q}[i-1,j-1] + f(i,j)\}$. When $P_i \neq Q_j$, then $f(i,j) = 1$, otherwise $f(i,j) = 0$. $L_{P,Q}[|P|, |Q|]$ is the Levenshtein distance between protein sequence P and protein sequence Q.

With Levenshtein distance of P and Q, the similarity between them can be calculated using the following formula:

$$\text{Sim}_{P,Q} = 1 - \frac{L_{P,Q}(|P|, |Q|)}{\max(|P|, |Q|)} \quad (4)$$

As with the method for generating drug structure similarity matrix, the following is the generated protein structure similarity matrix:

$$M^P = \begin{pmatrix} \text{Sim}(MF_{\Phi_1^P}, MF_{\Omega_1^P}) & \dots & \text{Sim}(MF_{\Phi_1^P}, MF_{\Omega_j^P}) \\ \vdots & \ddots & \vdots \\ \text{Sim}(MF_{\Phi_i^P}, MF_{\Omega_1^P}) & \dots & \text{Sim}(MF_{\Phi_i^P}, MF_{\Omega_j^P}) \end{pmatrix} \quad (5)$$

where the i-th row in M^P represents structural similarity feature vector V_i^{PSSF} of the i-th protein in inputting protein dataset. $\text{Sim}(MF_{\Phi_i^P}, MF_{\Omega_j^P})$ is the similarity score between the i-th protein in input dataset Φ^P and the j-th protein in benchmark dataset Ω^P .

- Generation of drug structural similarity vector (V^{DSSF}) and protein structural similarity vector (V^{PSSF})

Using the method introduced above, a drug structure similarity matrix and a protein structure similarity matrix can be generated based on the drugs and proteins in dataset respectively. Our method can implement generated drug (protein) structure similarity matrix when inputting individual drug SMILES /protein sequence and looks up V^{DSSF}/V^{PSSF} of inputting drug / protein in the matrix, as shown in Fig. 2. Similarity calculations are performed between drugs fed into DeepFusion and those in benchmark set to generate a drug structure similarity matrix. When an individual drug is fed into DeepFusion, the corresponding DSSF can be extracted in drug structure similarity matrix based only on the DrugBank ID of the individual drug. The generated V^{DSSF} and V^{PSSF} are then fed into convolutional neural network.

The dimensionality of the similarity features obtained at this point is large and some of the similarity information is redundant. If these features are used directly for feature fusion, this may result in inaccurate prediction. Therefore, we use CNN to extract feature from obtained drug and protein similarity feature with reduced dimensionality. Generating similarity-based interaction function F_{sim} :

$$I_{sim} = F_{sim}(CNN(V^{DSSF}), CNN(V^{PSSF})) \quad (6)$$

where $CNN()$ is three 1D convolutional neural network layers that extracts 1D feature vector. By F_{sim} , we denote the aggregation function for calculating similarity-based interaction, where we choose dot product as the aggregation function. Dimensionally reduced V^{DSSF} and V^{PSSF} represent a drug and protein respectively and dot product effectively maintains the structural integrity of drug and protein for the calculation of DTI at molecular level. Finally, we input the gained I_{sim} into IP-Module.

2.2.2. Extraction of local sub-structure feature

In order to get feature of each drug based on structural information of other molecules, we use structural similarity feature of molecules. This can be used to a great extent to capture unknown DTIs using known DTIs of other molecules. However, the occurrence of DTIs is influenced at the molecular substructure level, where interactions between drugs and protein sub-structures directly determine whether or not DTIs will occur. For this reason, we feed drug SMILES and protein sequence to the Sub-Channel of DeepFusion to extract drug and protein sub-structure information to better capture sub-structure interaction feature.

Frequent consecutive sub-sequence (FCS) mining algorithm [22] is used to extract sub-structure information from input molecules. After obtaining decomposed molecular sub-structure, we use transformer to enrich chemical connections of the sub-structure. The previous method [22] is used to predict DTIs after sub-structure information had been extracted using transformer. The generated interaction feature has a large dimensionality and are prone to the curse of dimensionality, which does not effectively represent interactions.

Let the drug SMILES to be D , and protein sequence be P . After the two sequences have been fed into sub-structure feature extraction channel, the decomposed sub-structure matrix $M_{dec}^D \in \mathbb{R}^{m \times s}$ and $M_{dec}^P \in \mathbb{R}^{n \times t}$ can be updated by FCS algorithm, in which m/n is actual size of drug/protein sub-structure and s/t is maximum length of drug/protein sub-structure. The calculation formula is as follows:

$$M_{dec}^D = FCS(D), M_{dec}^P = FCS(P) \quad (7)$$

where column $M_{dec_i}^D$ and $M_{dec_j}^P$ represents sub-structure between the i -th drug and the j -th protein. The presence or absence of sub-structure is indicated by one-hot vector.

Next, we need to calculate the content embedding and position embedding of sub-structure, and sum these two embeddings to obtain the final embeddings E^D and E^P . The final embedding is then fed into

transformer to better learn the sub-structural and chemical semantic information of input sequence. The specific formula is as follows:

$$E^D = Transformer_{Drug}\left(W_{cont}^D M_{dec_i}^D + W_{pos}^D H_i^D\right) \quad (8)$$

$$E^P = Transformer_{Protein}\left(W_{cont}^P M_{dec_j}^P + W_{pos}^P H_j^P\right) \quad (9)$$

where $Transformer()$ is a stacked 12-layer encoder based on self-attentive mechanism [23]. $W_{cont}^D \in \mathbb{R}^{l \times m}$ and $W_{cont}^P \in \mathbb{R}^{l \times n}$ are learnable dictionary lookup matrices used to generate content embeddings and position embeddings are generated using a lookup dictionary $W_{pos}^D \in \mathbb{R}^{l \times s}$ and $W_{pos}^P \in \mathbb{R}^{l \times t}$. And l is the size of the latent embedding of each sub-structure. $H_i^D \in \mathbb{R}^s$ and $H_j^P \in \mathbb{R}^t$ are one-hot vectors with value 1 at the i -th and j -th positions.

After obtaining E^D and E^P , we use a layer of Conv1D to dimensionally reduce them, again to prevent dimensionality curse. And generating sub-structure-based interaction by using function F_{sub} :

$$I_{sub} = F_{sub}(CNN(E^D), CNN(E^P)) \quad (10)$$

where $CNN()$ is a 1D convolutional neural network layer. Whereas, F_{sub} uses dot product to calculate sub-structure-based interaction. E^D and E^P are embedding vectors representing all sub-structures of drug and protein respectively. Dot product is selected to calculate the interactions between each sub-structure separately.

2.2.3. Interaction predicting module (IP-Module)

In DeepFusion, it extracts two features of interaction through Sim-Channel and Sub-Channel. After flattening I_{sim} and I_{sub} , we use function F on them to calculate interaction I by

$$I = F(Flatten(I_{sim}), Flatten(I_{sub})) \quad (11)$$

where function F uses concatenation for aggregation. Since I_{sim} and I_{sub} are interactions on the same level, the use of concatenation maintains the integrity of its own features.

Predicting network is made up of linear layer and ReLu function. The weights and bias parameters of whole network are optimized by binary classification loss. Interaction fusion feature I is fed to predicting network, which finally outputs a probability score P on the interaction by:

$$P = \text{Sigmoid}(WI + b) \quad (12)$$

where W and b is weight matrix and bias vector of the linear layer.

3. Results and discussion

It is used Rdkit to generate structural similarity matrices for drug and protein, Subword-nmt to construct sub-structure corpus of sequences and PyTorch to implement our proposed model. Some other necessary programming environments include Numpy, Pandas, Sklearn, Collections, Math, etc. Adam optimizer is applied with learning rate of 1e-4 and we set batch size to 64, epochs to 100 and dropout rate to 0.5. For experimental equipment, we use two Intel(R) Xeon(R) Gold 6226R 2.90 GHz CPUs and two NVIDIA Tesla V100 GPUs servers for our experiment.

3.1. Experiment design

Five data experiments are designed and conducted with the following objectives:

- (1) Evaluate the performance of DeepFusion in predicting DTI on different dataset sizes and negative ratios in data experiment 1.

Table 1

Description of drugs, proteins and positive and negative samples for the four sub-datasets of BIOSNAP and DAVIS dataset.

| Dataset | Drugs | Proteins | Positive Samples | Negative samples |
|---------|-------|----------|------------------|------------------|
| 100% | 4,510 | 2,181 | 13,741 | 13,742 |
| 70% | 4,404 | 2,173 | 9,618 | 9,618 |
| 50% | 4,140 | 2,154 | 6,864 | 6,865 |
| 30% | 3,502 | 2,030 | 4,121 | 4,121 |
| DAVIS | 68 | 379 | 2,666 | 9,597 |

- (2) Compare the robustness of DeepFusion with other state-of-the-art methods on different dataset sizes in data experiment 2.
- (3) Validate the necessity of using CNN for feature extraction between feature extraction channel and IP-Module in data experiment 3.
- (4) Analyze the contribution of structural similarity feature extraction channel and sub-structure feature extraction channel to DeepFusion in data experiment 4.
- (5) Verify the practical performance by molecular docking in case study.

3.2. Experimental setup

3.2.1. Dataset

We select here the MINER DTI dataset from BIOSNAP collection [26] for our data experiments. It contains 4,510 drugs and 2,181 proteins, and also includes 13,741 drug-target interaction pairs from DrugBank [27]. Since BIOSNAP contains only positive DTI pairs, we expand negative samples by the number of positive samples in order to balance data. In this work, negative samples are defined as drug-target pair has no interaction or is unknown. We also introduce DAVIS dataset [28] as a highly unbalanced dataset to evaluate the robustness of our model, and the data details are presented in Table 1. For benchmark set of Drug/protein, we remove the duplicate drug and protein data from BIOSNAP and DAVIS separately to obtain baseline drug dataset and baseline protein dataset with no duplicate data. We then use them separately to generate drug/protein structural similarity matrix. For subsequent experiments, we randomly split BIOSNAP dataset into four sub-datasets according to 100%, 70%, 50% and 30% of the dataset size, as shown in Table 1. Also, we select 70% of the dataset to be used as training set, 10% as validation set and 20% as test set.

3.2.2. Evaluation metrics

We evaluate the performance of model using metrics that are commonly used in classification experiments, including: ROC-AUC (Area under the receiver operating characteristic curve), PR-AUC (Area under the precision-recall curve), accuracy and F1 score. The smoother ROC-AUC and PR-AUC curves and larger AUC values indicate better model performance, and they also ignore the effect of threshold selection, which in practice is often determined by a priori probabilities or artificially.

3.3. Baselines

Morgan-CNN [15] inputs Morgan Fingerprint of drug into a multi-layer perceptron. Protein sequences are decomposed into individual characters and represented in embeddings, which are then fed into a multi-layer 1D CNN.

MPNN-CNN uses message passing neural network [29] to extract drug feature. It can extract embedding vector to molecular graph-level by embedding vector for each atom and edge. A multi-layer perceptron decoder is used to predict DTIs.

DeepConv-DTI [16] takes ECFP4 of drug as input and uses a fully connected layer to extract drug features. CNN and global maximum

Table 2

DeepFusion achieved the best prediction performance on all datasets (average of five experiments).

| Dataset | Method | ROC-AUC | PR-AUC | Accuracy | F1 |
|---------|----------------|---------|--------|----------|-------|
| 100% | Morgan-CNN | 0.876 | 0.885 | 0.800 | 0.795 |
| | MPNN-CNN | 0.873 | 0.886 | 0.798 | 0.795 |
| | DeepConv-DTI | 0.876 | 0.890 | 0.803 | 0.796 |
| | GraphDTA | 0.890 | 0.898 | 0.811 | 0.802 |
| | TransformerCPI | 0.892 | 0.903 | 0.808 | 0.802 |
| | MolTrans | 0.882 | 0.894 | 0.795 | 0.802 |
| | DeepFusion | 0.901 | 0.910 | 0.830 | 0.827 |
| 70% | Morgan-CNN | 0.854 | 0.861 | 0.781 | 0.792 |
| | MPNN-CNN | 0.853 | 0.864 | 0.783 | 0.777 |
| | DeepConv-DTI | 0.856 | 0.873 | 0.782 | 0.784 |
| | GraphDTA | 0.869 | 0.884 | 0.781 | 0.794 |
| | TransformerCPI | 0.874 | 0.883 | 0.790 | 0.797 |
| | MolTrans | 0.872 | 0.876 | 0.795 | 0.800 |
| | DeepFusion | 0.877 | 0.888 | 0.796 | 0.802 |
| 50% | Morgan-CNN | 0.822 | 0.838 | 0.694 | 0.743 |
| | MPNN-CNN | 0.841 | 0.858 | 0.755 | 0.758 |
| | DeepConv-DTI | 0.815 | 0.837 | 0.694 | 0.734 |
| | GraphDTA | 0.840 | 0.862 | 0.748 | 0.763 |
| | TransformerCPI | 0.846 | 0.850 | 0.768 | 0.755 |
| | MolTrans | 0.848 | 0.835 | 0.766 | 0.770 |
| | DeepFusion | 0.859 | 0.874 | 0.771 | 0.776 |
| 30% | Morgan-CNN | 0.778 | 0.812 | 0.698 | 0.726 |
| | MPNN-CNN | 0.819 | 0.850 | 0.729 | 0.695 |
| | DeepConv-DTI | 0.777 | 0.812 | 0.718 | 0.702 |
| | GraphDTA | 0.841 | 0.859 | 0.753 | 0.733 |
| | TransformerCPI | 0.826 | 0.857 | 0.729 | 0.755 |
| | MolTrans | 0.817 | 0.831 | 0.732 | 0.737 |
| | DeepFusion | 0.836 | 0.870 | 0.765 | 0.760 |
| DAVIS | Morgan-CNN | 0.884 | 0.293 | 0.794 | 0.291 |
| | MPNN-CNN | 0.767 | 0.167 | 0.771 | 0.220 |
| | DeepConv-DTI | 0.880 | 0.340 | 0.858 | 0.337 |
| | GraphDTA | 0.874 | 0.289 | 0.677 | 0.214 |
| | TransformerCPI | 0.874 | 0.277 | 0.745 | 0.254 |
| | MolTrans | 0.907 | 0.375 | 0.849 | 0.367 |
| | DeepFusion | 0.911 | 0.402 | 0.837 | 0.388 |

pooling layers are used to extract local models of different lengths in protein sequence.

GraphDTA [18] applies GNN and CNN to drug and protein representation, respectively. In our experiments, we added sigmoid activation function to the last layer for DTI prediction.

TransformerCPI [19] is based on transformer architecture, a model that generates atomic representations and protein sequence representations from CNN and GCN for drug and protein sequences, respectively. Interaction feature are obtained through transformer's decoder and interaction probabilities are output using a linear layer. Moltrans [22] uses Frequent Consecutive Sub-sequence mining to extract sub-structures of SMILES and protein sequence, and transformer for enhanced representation of sub-structures. The model is by far the most advanced for the DTI prediction task and we use the hyper-parameters in the original article for experiment.

3.4. Data experiment 1

We evaluate performance of DeepFusion on four sub-datasets of BIOSNAP. It is obtained from Table 2 that DeepFusion has the best overall performance on all sub-datasets. Specifically, it is the highest in all evaluation metrics on 100%, 70% and 50% sub-dataset. On 70% sub-dataset, it achieves ROC-AUC and PR-AUC are 0.877 and 0.888 respectively, which is close to the performance of some baseline methods. On 30% sub-dataset, DeepFusion outperform the superior baseline method Moltrans by 1.9% and 3.9%, and outperform

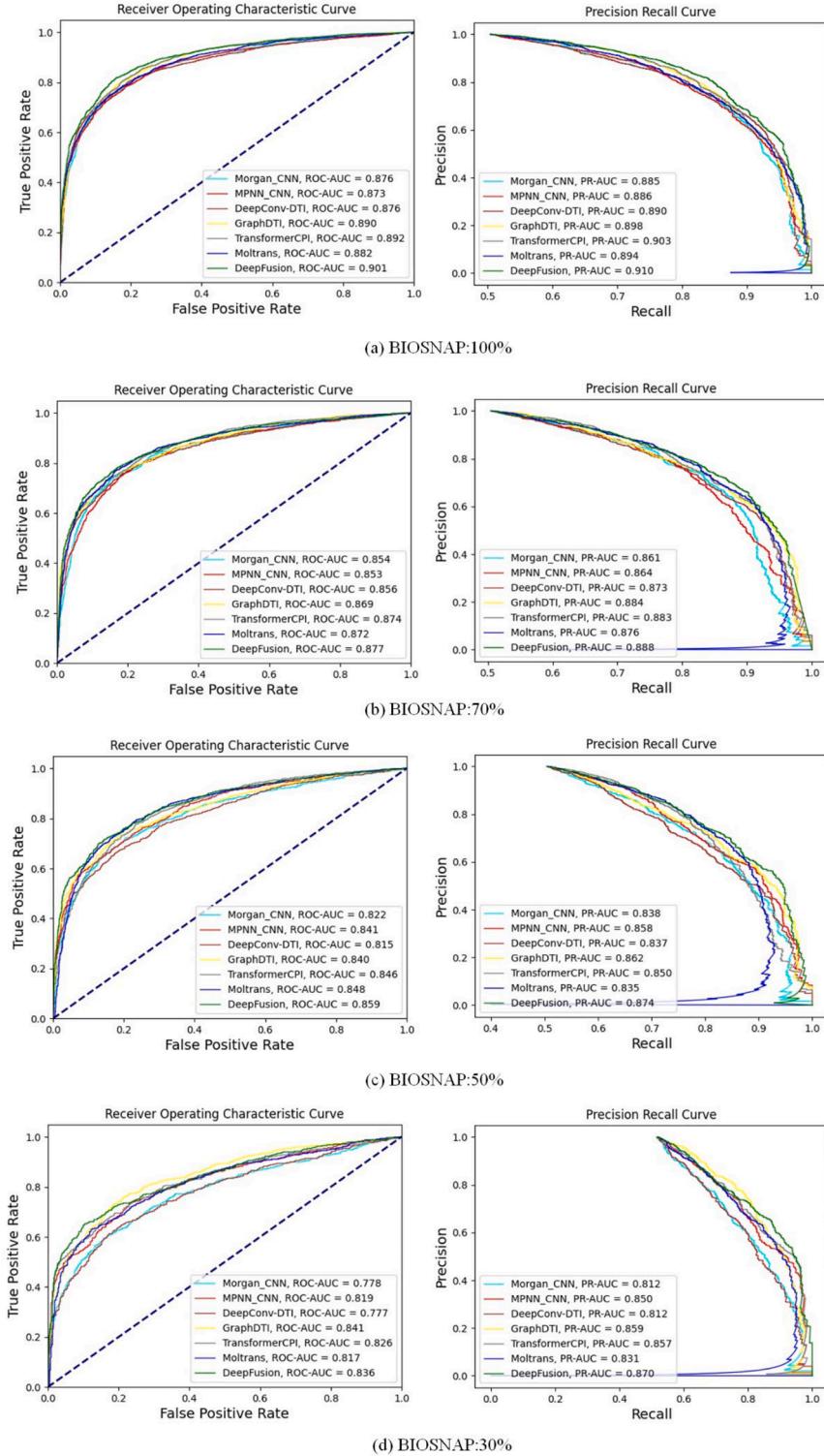


Fig. 3. The ROC-AUC of DeepFusion performs best on all sub-datasets.

DeepConv-DTI by 5.9% and 5.8% for ROC-AUC and PR-AUC. On DAVIS dataset, a highly unbalanced positive and negative sample distribution, DeepFusion achieves the highest performance in all metrics except accuracy, indicating that our model is equally good at predicting on unbalanced datasets with very good robustness.

It is worth noticing that although MPNN-CNN performs poorly on large-scale dataset, it learns well on small-scale dataset. We attribute this to the fact that MPNN exploits information at multiple levels such as atomic properties and molecular structure.

3.5. Data experiment 2

Four sub-datasets of BIOSNAP are used in this experiment to evaluate the robustness of DeepFusion and baseline methods. These four sub-datasets have a relatively large difference in size and model settings are not biased when making prediction. We plot ROC and PR curves by the actual labels of DTIs and predicted values of each model, as shown in Fig. 3. During the reduction of the dataset size, ROC curve and PR curve of DeepFusion are very smooth and both ROC-AUC and PR-AUC are the

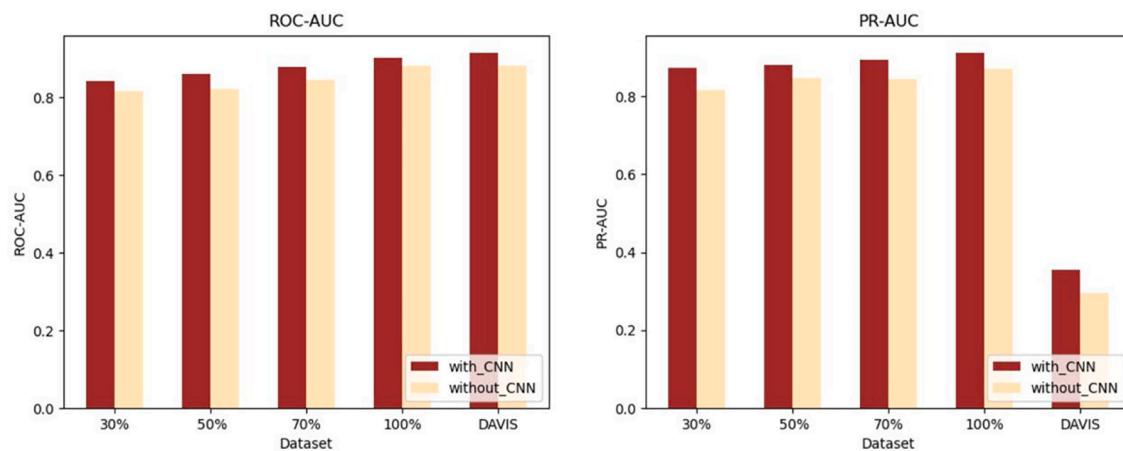


Fig. 4. Model with CNN perform better than without on all datasets.

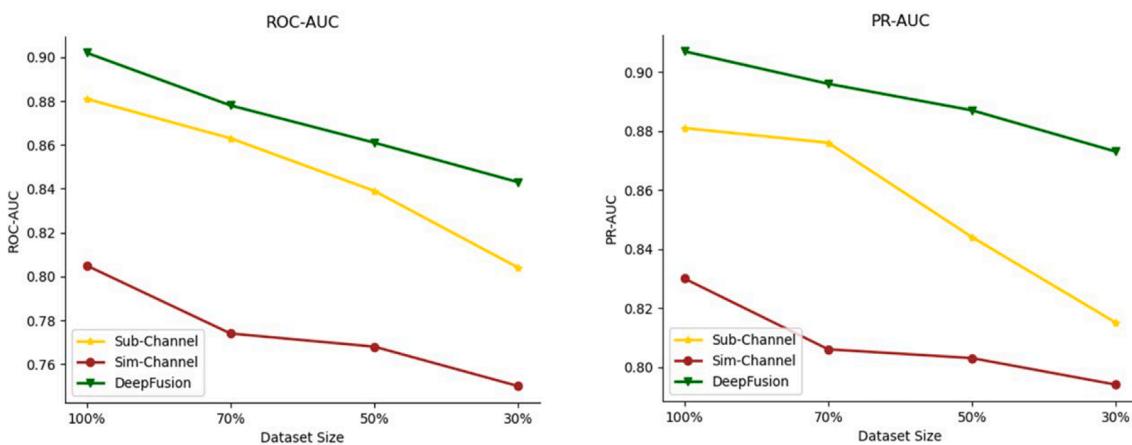


Fig. 5. Sub-Channel prediction performs better and Sim-Channel prediction performance is less affected by dataset size.

best.

In contrast, the other methods show larger fluctuations in both ROC and PR curve and huge decreases in ROC-AUC and PR-AUC. Thus DeepFusion has the best robustness and the best predictive performance compared to the baseline method. In practice, when predicting interaction of a certain target protein with drugs, the data scale is not too large and DeepFusion allows for better prediction.

Table 3
DeepFusion makes new prediction for 30% sub-dataset (in no particular order).

| Protein ID | Drug ID | Actual Label | Predictive Label | Evidences |
|------------|---------|--------------|------------------|-----------|
| P78527 | DB00201 | 1 | 1 | [30] |
| P08684 | DB00295 | 1 | 1 | [31–33] |
| P10635 | DB00334 | 1 | 1 | [34] |
| Q9UBM7 | DB00157 | 1 | 1 | [35,36] |
| P48775 | DB00150 | 1 | 1 | [37] |
| P10632 | DB00897 | 1 | 1 | [38] |
| P47870 | DB00898 | 1 | 1 | [39,40] |
| P10632 | DB00688 | 1 | 1 | [41,42] |
| O95168 | DB00157 | 1 | 1 | [35] |
| O15067 | DB00142 | 1 | 1 | [35] |
| P08684 | DB00540 | 1 | 1 | [43,44] |
| P08684 | DB00688 | 1 | 1 | [40,42] |
| P02679 | DB00157 | 0 | 1 | Unknow |
| P13498 | DB00786 | 0 | 1 | Unknow |

3.6. Data experiment 3

To validate the necessity of using CNN between feature extraction channel and IP-Module, we need to perform experiments on models with and without CNN. We conduct experiments on four sub-datasets and DAVIS. The results are shown in Fig. 4. It can be seen that the model using CNN after feature extraction channel performs better than the model without CNN. During the period of conducting experiments, we observed that dimensionality of features extracted through Sim-channel and Sub-channel are very large. Thus, implementation of CNN at this point can effectively remove redundant information and obtain feature with stronger relevance, which improves the final prediction performance.

3.7. Data experiment 4

There are Sim-Channel and Sub-Channel to extract features of drug and protein. To compare feature extraction capabilities of the two channels, we use one of the channels in combination with interaction feature fusion module for DTIs prediction respectively, and the results obtained are shown in Fig. 5.

In terms of evaluation metrics, the prediction performance using Sub-Channel alone is better than using Sim-Channel alone. However, as the size of the dataset decreases, prediction performance drops dramatically while Sim-channel keeps a high stability. However, when the dataset size was reduced from 70% to 30%, ROC-AUC of sub-

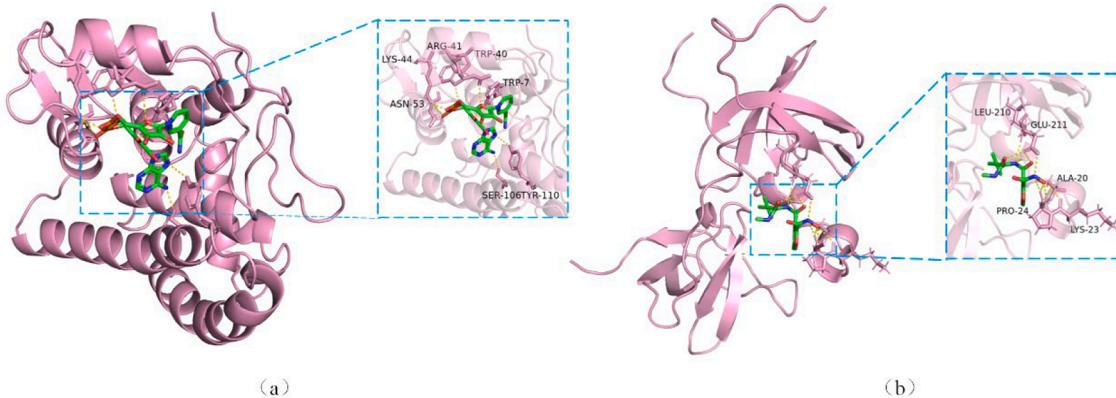


Fig. 6. Two predicted DTI pairs by DeepFusion.

structure channel decreased by 7.3% and PR-AUC by 8.8%. However, the reduction in dataset size has very little impact on Sim-Channel, with ROC-AUC dropping by only 3.2% and PR-AUC dropping by 1.5% when the dataset size is reduced from 70% to 30%. This is attributed to the fact that Sim-Channel can use features of other molecules to determine their own features, thus extracting more features on limited data. DeepFusion has the advantage of strong predictive power in Sub-Channel and inherits the advantage of strong feature extraction in Sim-Channel, resulting in better prediction performance with less impact from dataset size reduction.

3.8. Case study

It is verified that DeepFusion has a high confidence level, which is useful for practical applications, so we do prediction experiments on 30% sub-dataset of BIOSNAP. Twelve prediction scores of the top fourteen DTI pairs are confirmed by previous literature or included in DrugBank, and two are newly predicted, as shown in Table 3 (in no particular order, as the top 14 prediction scores are all 1).

The prediction score is the unthresholded probability score predicted by DeepFusion. If the value is 1, it indicates that DeepFusion believes that the drug-protein pair has interaction. Thus, P02679 and DB00157, P13498 and DB00786 are likely to be drug-protein pairs with interactions that have not yet been identified. We use PDB entries for molecular docking on Autodock Vina [45] and visualise the hydrogen bonds formed by some of these amino acids and drug molecules. Molecular docking and hydrogen bonding coloring between 1DUG (PDB entry of P02679) and DB00157, as shown in Fig. 6(a). Molecular docking and hydrogen bonding coloring between 1WLP (PDB entry of P13498) and DB00786, as shown in Fig. 6(b).

4. Conclusions

In this paper, we present DeepFusion framework, which can effectively extract molecular structural similarity feature and sub-structure feature, which are fed into interaction feature fusion module to encode interaction feature. We conduct experiments using real-world dataset and compare them with four currently more advanced methods. The results show that DeepFusion has excellent predictive performance on large-scale, small-scale and unbalanced datasets. We use DeepFusion to predict two new DTI pairs in BIOSNAP sub-dataset and validate them through molecular docking method, demonstrating that our prediction is of some practical relevance and application. DeepFusion also has some limitations, in this paper the similarities of drugs and proteins are calculated based on 2D and 1D structural information, respectively, which may lose some information. In the future we will collect 3D structural information of drugs and proteins and use it to construct 3D model.

Funding

This work was supported by Natural Science Foundation of China (Grant Nos. 61873280, 61672033, 61672248, 61972416), Taishan Scholarship (tsqn201812029), Natural Science Foundation of Shandong Province (No. ZR2019MF012), Foundation of Science and Technology Development of Jinan (201907116) and Fundamental Research Funds for the Central Universities (18CX02152A, 19CX05003A-6).

CRediT authorship contribution statement

Tao Song: Conceptualization, Investigation, Resources, Writing – original draft, Project administration. **Xudong Zhang:** Methodology, Software, Validation, Visualization. **Mao Ding:** Validation, Supervision. **Alfonso Rodriguez-Paton:** Validation, Formal analysis, Writing – review & editing, Funding acquisition. **Shudong Wang:** Conceptualization. **Gan Wang:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. Knowles, G. Gromi, A guide to drug discovery: target selection in drug discovery, *Nat. Rev. Drug Discov.* 2 (1) (2003) 63–69, <https://doi.org/10.1038/nrd986>.
- [2] N. Novac, Challenges and opportunities of drug repositioning, *Trends Pharmacol. Sci.* 34 (5) (2013) 267–272, <https://doi.org/10.1016/j.tips.2013.03.004>.
- [3] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, H. Lu, Deep-learning-based drug-target interaction prediction, *J. Proteome Res.* 16 (4) (2017) 1401–1409, <https://doi.org/10.1021/acs.jproteome.6b00618>.
- [4] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, *Mol. Inf.* 29 (6–7) (2010) 476–488, <https://doi.org/10.1002/minf.201000061>.
- [5] F.J. Prado-Prado, F. Borges, L.G. Perez-Montoto, H. González-Díaz, Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species, *Eur. J. Med. Chem.* 44 (10) (2009) 4051–4056.
- [6] F.J. Prado-Prado, E. Uriarte, F. Borges, H. González-Díaz, Multi-target spectral moments for QSAR and complex networks study of antibacterial drugs, *Eur. J. Med. Chem.* 44 (11) (2009) 4516–4521, <https://doi.org/10.1016/j.ejmech.2009.06.018>.
- [7] L.B. Salum, A.D. Andricopulo, Fragment-based QSAR: perspectives in drug design, *Mol. Diversity* 13 (3) (2009) 277–285, <https://doi.org/10.1007/s11030-009-9112-5>.
- [8] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat. Rev. Drug Discovery* 3 (11) (2004) 935–949, <https://doi.org/10.1038/nrd1549>.
- [9] T. Song, S. Wang, D. Liu, M. Ding, Z. Du, Y. Zhong, A. Rodri Guez-Patón, SE-OnionNet: A convolutional neural network for protein-ligand binding affinity prediction, *Frontiers in Genetics* 11 (2020) 1805. <https://doi.org/10.3389/fgene.2020.607824>.
- [10] D.B. Cao, L.X. Zhang, G.S. Tan, Z. Xiang, W.B. Zeng, Q.S. Xu, A.F. Chen, Computational prediction of drug-target interactions using chemical, biological,

- and network features, *Molecular informatics* 33(10) (2014) 669–681. <https://doi.org/10.1002/minf.201400009>.
- [11] E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, Comparison of support vector machine and artificial neural network systems for drug/nondrug classification, *J. Chem. Inf. Comput. Sci.* 43 (6) (2003) 1882–1889, <https://doi.org/10.1021/ci0341161>.
- [12] A. Mayr, G. Klambauer, T. Unterthiner, M. Steijaert, J.K. Wegner, H. Ceulemans, D.-A. Clevert, S. Hochreiter, Large-scale comparison of machine learning methods for drug target prediction on ChEMBL, *Chem. Sci.* 9 (24) (2018) 5441–5451, <https://doi.org/10.1039/c8sc00148k>.
- [13] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug-target binding affinity prediction, *Bioinformatics* 34 (17) (2018) i821–i829, <https://doi.org/10.1093/bioinformatics/bty593>.
- [14] S. Wang, Z. Du, M. Ding, R. Zhao, A. Rodriguez-Paton, T. Song, LDCNN-DTI: a novel light deep convolutional neural network for drug-target interaction predictions, *IEEE* (2020) 1132–1136, <https://doi.org/10.1109/BIBM49941.2020.9313585>.
- [15] K. Huang, T. Fu, L.M. Glass, M. Zitnik, C. Xiao, J. Sun, DeepPurpose: a deep learning library for drug-target interaction prediction, *Bioinformatics* 36 (22–23) (2020) 5545–5547. <https://doi.org/10.1093/bioinformatics/btaa1005>.
- [16] I. Lee, J. Keum, H. Nam, J.M. Briggs, DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.* 15 (6) (2019) e1007129, <https://doi.org/10.1371/journal.pcbi.1007129>.
- [17] G. Taherzadeh, Y. Zhou, A.W.-C. Liew, Y. Yang, Structure-based prediction of protein-peptide binding regions using Random Forest, *Bioinformatics* 34 (3) (2018) 477–484. <https://doi.org/10.1093/bioinformatics/btx614>.
- [18] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: Predicting drug-target binding affinity with graph neural networks, *Bioinformatics* 37 (8) (2021) 1140–1147. <https://doi.org/10.1093/bioinformatics/btaa921>.
- [19] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, M. Zheng, TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments, *Bioinformatics* 36 (16) (2020) 4406–4414. <https://doi.org/10.1093/bioinformatics/btaa524>.
- [20] F. Rayhan, S. Ahmed, Z. Mousavian, D.M. Farid, S. Shatabda, FRnet-DTI: deep convolutional neural network for drug-target interaction prediction, *Heliyon* 6 (3) (2020) e03444, <https://doi.org/10.1016/j.heliyon.2020.e03444>.
- [21] T. Song, Y. Zhong, M. Ding, R. Zhao, Q. Tian, Z. Du, D. Liu, J. Liu, Y. Deng, Repositioning molecules of Chinese medicine to targets of SARS-CoV-2 by deep learning method, *IEEE* (2020) 2306–2312, <https://doi.org/10.1109/BIBM49941.2020.9313151>.
- [22] K. Huang, C. Xiao, L.M. Glass, J. Sun, MolTrans: Molecular Interaction Transformer for drug-target interaction prediction, *Bioinformatics* 37 (6) (2021) 830–836. <https://doi.org/10.1093/bioinformatics/btaa880>.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, 2017, pp. 5998–6008.
- [24] M. Hattori, Y. Okuno, S. Goto, M. Kanehisa, Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways, *J. Am. Chem. Soc.* 125 (39) (2003) 11853–11865, <https://doi.org/10.1021/ja036030u>.
- [25] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Union*, 1966, pp. 707–710.
- [26] M. Zitnik, R. Sosi, S. Maheshwari, J. Leskovec, BioSNAP Datasets: Stanford biomedical network dataset collection., 2018.
- [27] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M.J. N.a.r. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, 36(suppl_1) (2008) D901–D906. <https://doi.org/10.1093/nar/gkm958>.
- [28] M.I. Davis, J.P. Hunt, S. Herrgard, P. Ciceri, L.M. Wodicka, G. Pallares, M. Hocker, D.K. Treiber, P.P.J.N.B. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, 29(11) (2011) 1046–1051.
- [29] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Neural message passing for quantum chemistry, *International conference on machine learning*, PMLR, 2017, pp. 1263–1272. <https://doi.org/10.138/nbt.1990>.
- [30] L.C. Foukas, N. Daniele, C. Ktori, K.E. Anderson, J. Jensen, P.R.J.J.O.B.C. Shepherd, Direct effects of caffeine and theophylline on p110 α and other phosphoinositide 3-kinases: differential effects on lipid kinase and protein kinase activities, 277(40) (2002) 37124–37130. <https://doi.org/10.1074/jbc.M202101200>.
- [31] E.B. Daily, C.L. Aquilante, *Cytochrome P450 2C8 pharmacogenetics: a review of clinical studies*, *Pharmacogenomics* 10 (9) (2009) 1489–1510.
- [32] D. Projean, P.-E. Morin, T. Tu, J.X. Ducharme, Identification of CYP3A4 and CYP2C8 as the major cytochrome P450 s responsible for morphine N-demethylation in human liver microsomes, 33(8) (2003) 841–854. <https://doi.org/10.1080/0049825031000121608>.
- [33] S. Takeda, Y. Ishii, P.I. Mackenzie, K. Nagata, Y. Yamazoe, K. Oguri, H. Yamada, Modulation of UDP-glucuronosyltransferase 2B7 function by cytochrome P450s in Vitro: differential effects of CYP1A2, CYP2C9 and CYP3A4, *Biol. Pharm. Bull.* 28 (10) (2005) 2026–2027.
- [34] J.T. Callaghan, R.F. Bergstrom, L.R. Ptak, C.M.J.C.P. Beasley, Olanzapine. Pharmacokinetic and pharmacodynamic profile, 37(3) (1999) 177. <https://doi.org/10.2165/00003088-199937030-00001>.
- [35] P. Imming, C. Sinning, A.J.N.r.D.D. Meyer, Drugs, their targets and the nature and number of drug targets, 5(10) (2006) 821–834. <https://doi.org/10.1038/nrd2132>.
- [36] J.P. Overington, B. Al-Lazikani, A.L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* 5 (12) (2006) 993–996.
- [37] S.A. Rafice, N. Chauhan, I. Efimov, J. Basran, E.L.J.B.S.T. Raven, Oxidation of L-tryptophan in biology: a comparison between tryptophan 2, 3-dioxigenase and indoleamine 2, 3-dioxigenase, 37(2) (2009) 408–412. <https://doi.org/10.1042/BST0370408>.
- [38] C.E. Ong, S. Coulter, D.J. Birkett, C.R. Bhasker, J.O.J.B.j.o.c.p. Miners, The xenobiotic inhibitor profile of cytochrome P4502C8, 50(6) (2000) 573–580. <https://doi.org/10.1046/j.1365-2125.2000.00316.x>.
- [39] V. Santhakumar, M. Wallner, T.S.J.A. Otis, Ethanol acts directly on extrasynaptic subtypes of GABA_A receptors to increase tonic inhibition, 41(3) (2007) 211–221. <https://doi.org/10.1016/j.alcohol.2007.04.011>.
- [40] M.J.J.o.p. Davies, neuroscience, The role of GABA_A receptors in mediating the effects of alcohol in the central nervous system, (2003). [https://doi.org/10.1016/S0022-3956\(03\)00051-7](https://doi.org/10.1016/S0022-3956(03)00051-7).
- [41] N. Picard, T. Cresteil, A. Prémaud, P.J.T.d.m. Marquet, Characterization of a phase 1 metabolite of mycophenolic acid produced by CYP3A4/5, 26(6) (2004) 600–608. <https://doi.org/10.1097/00007691-200412000-00004>.
- [42] V. Lamba, K. Sangkuhl, K. Sanghavi, A. Fish, R.B. Altman, T.E.J.P. Klein, genomics, PharmGKB summary: mycophenolic acid pathway, 24(1) (2014) 73. <https://doi.org/10.1097/FPC.0000000000000010>.
- [43] B.O. Wen, L.i. Ma, M. Zhu, Bioactivation of the tricyclic antidepressant amitriptyline and its metabolite nortriptyline to arene oxide intermediates in human liver microsomes and recombinant P450s, *Chem. Biol. Interact.* 173 (1) (2008) 59–67.
- [44] K. Venkatakrishnan, L.L. von Moltke, D.J.J.T.J.o.C.P. Greenblatt, Nortriptyline E-10-hydroxylation in vitro is mediated by human CYP2D6 (high affinity) and CYP3A4 (low affinity): implications for interactions with enzyme-inducing drugs, 39(6) (1999) 567–577. <https://doi.org/10.1177/00912709922008173>.
- [45] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* (2009), <https://doi.org/10.1002/jcc.21334>.