

Predicting drug–target binding affinity with cross-scale graph contrastive learning

Jingru Wang, Yihang Xiao, Xuequn Shang and Jiajie Peng

Corresponding author. Jiajie Peng, E-mail: jiajiepeng@nwpu.edu.cn

Abstract

Identifying the binding affinity between a drug and its target is essential in drug discovery and repurposing. Numerous computational approaches have been proposed for understanding these interactions. However, most existing methods only utilize either the molecular structure information of drugs and targets or the interaction information of drug–target bipartite networks. They may fail to combine the molecule-scale and network-scale features to obtain high-quality representations. In this study, we propose CSCo-DTA, a novel cross-scale graph contrastive learning approach for drug–target binding affinity prediction. The proposed model combines features learned from the molecular scale and the network scale to capture information from both local and global perspectives. We conducted experiments on two benchmark datasets, and the proposed model outperformed existing state-of-art methods. The ablation experiment demonstrated the significance and efficacy of multi-scale features and cross-scale contrastive learning modules in improving the prediction performance. Moreover, we applied the CSCo-DTA to predict the novel potential targets for Erlotinib and validated the predicted targets with the molecular docking analysis.

Keywords: drug discovery; drug–target binding affinity; cross-scale; graph contrastive learning

INTRODUCTION

It takes an average of over a decade for a drug to get from the initial research and development stage to the commercialization stage [1–3], and the estimated cost for this process exceeds \$1 billion [4]. Developing new drug involves several stages, including target identification [5, 6], drug lead discovery [7, 8], preclinical development [9] and clinical trials [10]. A report reveals most drug development failures in clinical trials resulting from insufficient target validation [11]. The result highlights the significance and necessity of accurate drug screening for specific targets.

As a result, identifying interactions between drug molecules and receptor targets plays a crucial role in drug discovery. The goal is to identify compounds that can bind to disease targets with an activating or inhibiting effect, akin to matching a key to a lock [12]. These intermolecular interactions between compounds and targets occur within the pocket-like region of the protein, known as the binding site, where specific amino acid residues play a crucial role [13]. Compounds can engage with the protein targets in various orientations, influencing the efficacy and affinity of the binding process [14]. The binding affinity, along with the binding site and the binding conformation, form an interconnected system that influences the effectiveness of the interaction. Evaluating the binding affinity provides a key metric

to assess the strength of interaction between a drug and its target. Experimentally determined parameters such as dissociation constant, inhibition constant and half-maximal inhibitory concentration are often used to indicate the binding strength between drugs and targets [15, 16]. However, determining these values by wet experiments is typically time-consuming and costly.

Over the past decades, computer-aided drug design technologies have had a revolutionary impact on the field of drug discovery. Molecular docking and dynamics simulation methods are frequently used to identify and predict potential drug targets, allowing researchers to accelerate the drug discovery and development process [17–19]. These methods can predict binding affinity by considering the geometric structural complementarity and binding free energy between ligand-acceptor molecules. These methods usually rely on the three-dimensional structures of the complexes. Although the tools like AlphaFold have enhanced the protein structure prediction, obtaining the dynamic protein structure and structure of protein complex is still challenging [20]. This limits the power of molecular docking and dynamics simulation methods.

Recent years have seen the emergence of artificial intelligence as a promising tool for drug–target prediction using machine learning and deep learning techniques. Unlike traditional computer-aided methods, these approaches do not strictly require

Jingru Wang is a doctoral student in the School of Computer Science at Northwestern Polytechnical University. Her research interests include computational biology and machine learning.

Yihang Xiao is a master student in the School of Computer Science at Northwestern Polytechnical University. His research interests include computational biology and machine learning.

Xuequn Shang is a professor in the School of Computer Science at Northwestern Polytechnical University. Her research interests include data mining and computational biology.

Jiajie Peng is an associate professor in the School of Computer Science at Northwestern Polytechnical University. His research interests include computational biology and machine learning.

Received: July 3, 2023. **Revised:** December 4, 2023. **Accepted:** December 7, 2023

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

the 3D structure of the ligand–receptor complex as input. Many studies have focused on binary drug–target interaction (DTI) prediction, and have achieved good results [21–27]. However, such type of prediction task may not adequately reflect the binding capacity of a drug and its target, nor accurately represent the efficacy of drug. As a result, more attention has been given to drug–target binding affinity (DTA) prediction recently, which can be defined as a regression task. Quite a few methods have been developed to address this task. For instance, Li *et al.* [28] proposed a docking scoring function based on a random forest model to predict affinity values between drugs and targets, and they found that low-quality data can also help improve the predictive performance of the machine-learning model. Pahikkala *et al.* [29] applied the regularized least squares method to predict the continuous DTA value. He *et al.* [30] presented a gradient-boosting tree model named SimBoost, which utilized informative features from the drug and target similarities to achieve the purpose of DTA prediction. And some methods use the sequence information of drugs and targets. For example, Öztürk *et al.* [31] employed double-independent convolutional neural network (CNN) blocks to extract features of drugs and targets from SMILES strings and protein sequences, respectively, and used learned representations for DTA prediction through a fully connected layer. Lin *et al.* [32] proposed a novel deep learning-based model named DeepGS. The model employs a CNN to learn representations of the target protein sequence and uses a graph attention network and a bidirectional gating recurrent unit to extract two types of drug features. These features are then concatenated with the protein sequence features to predict the binding affinity score. Li *et al.* [33] introduced the atom (in the compound)–residue (in the protein) pairwise interaction matrix to consider the non-covalent connections and predict the DTA. Zeng *et al.* [34] presented MATT_DTI, which further used a self-attention block to encode the interaction information between drug representations and protein representations. Yuan *et al.* [35] proposed a multi-head linear attention mechanism to aggregate information for drugs and targets and applied knowledge distillation in the training stage. In practice, each drug molecule can be modeled as a graph with atoms as nodes and bonds as edges. With the recent advancements in graph neural networks, feature extraction for graph-structured data has become more convenient and effective. This has led to the development of various graph-based models for drug–target affinity prediction. Nguyen *et al.* [36] represented each drug as a molecular graph and introduced graph neural networks to learn the drug feature representations, and the target feature representations are extracted through CNN by capturing sequence information. Zhang *et al.* [37] presented SAG-DTA that introduced self-attention mechanisms when learning the representations of the molecular drug graph. However, these methods still treated the target proteins as sequential data for processing. Jiang *et al.* [38] utilized protein structure prediction (Pcons4) to construct the protein contact map, and subsequently employed graph neural networks to learn features on the protein contact map and the molecular drug map. Most of the previous studies have utilized deep neural networks to learn feature representations of protein and drug molecules, which are then fed into a multilayer perceptron (MLP) for prediction. This framework is currently the foundation of machine learning models for predicting DTA. However, these methods only utilized the molecular structure information of drugs and proteins, with DTA information only appearing in the model training stage as the supervised task label. Existing methods did not consider the topology information of the drug–target network, which has proven to be useful in existing DTI prediction studies [21, 27].

These studies are achieved based on a drug–target network. In this network, drugs and targets are represented as nodes, and the connections between them are considered as edges, which correspond to the observed interactions between specific drugs and targets. Adopting this network-based approach enables us to predict unknown DTIs and understand the underlying mechanisms governing drug–target association [39, 40]. Molecular properties and drug–target network topology represent features of different scales. However, effectively capturing and integrating features from multiple scales remain a significant challenge.

To improve the performance of drug–target affinity prediction, we propose a novel **cross-scale graph contrastive learning-based drug-target affinity prediction model**, named CSCo-DTA, which differs from previous methods that primarily focus on using molecular features alone for DTA prediction. CSCo-DTA considers both network-scale feature and molecule-scale feature through a graph contrastive learning-based framework. CSCo-DTA consists of four main components, including molecule-scale feature extraction, network-scale feature extraction, contrastive learning across the molecule and network scales and DTA prediction. The proposed method specifically extracts network-scale features of drugs and proteins from the DTI network, as well as molecule-scale features of drugs and proteins from the drug and target molecule network, via graph neural networks. Next, using the feature representations of molecule-scale and network-scale, CSCo-DTA explores the consistency between the two perspectives through contrastive learning. Finally, an objective with a multi-task training strategy is designed to jointly optimize the representations of drugs and targets, as well as to predict DTA. We compare our method with existing methods on two benchmark datasets. The results demonstrate that CSCo-DTA performs the best on both datasets. Furthermore, we investigate the impact of multi-scale features of drugs and targets, as well as using cross-scale contrastive learning modules on the predictive performance of the model through the ablation experiment. The results demonstrate that these components significantly improve the accuracy of DTA prediction. Moreover, we conduct a case study on the potential targets of Erlotinib. The results indicate that our model can effectively predict the binding affinity of DTIs, which could have important implications for accelerating the drug discovery process. These findings highlight the potential of our approach to contribute to the development of more effective and targeted therapies for various diseases.

METHODS

In this section, we first formulate the DTA prediction task as a regression problem and then propose a **cross-scale graph contrastive learning-based drug-target affinity prediction model** (CSCo-DTA). CSCo-DTA leverages graph neural network encoders to extract both molecule-scale and network-scale features, and maximizes the mutual information between these two scale features via a cross-scale graph contrastive learning component. Figure 1 illustrates the framework of the proposed model, which comprises four parts: the molecule-scale graph representation learning component (Figure 1A,C), the network-scale graph representation learning component (Figure 1B), the cross-scale graph contrastive component (Figure 1D) and the drug–target affinity prediction component (Figure 1E).

Preliminaries

Before introducing the details of our method, we first formulate the DTA prediction as a regression problem and give the necessary

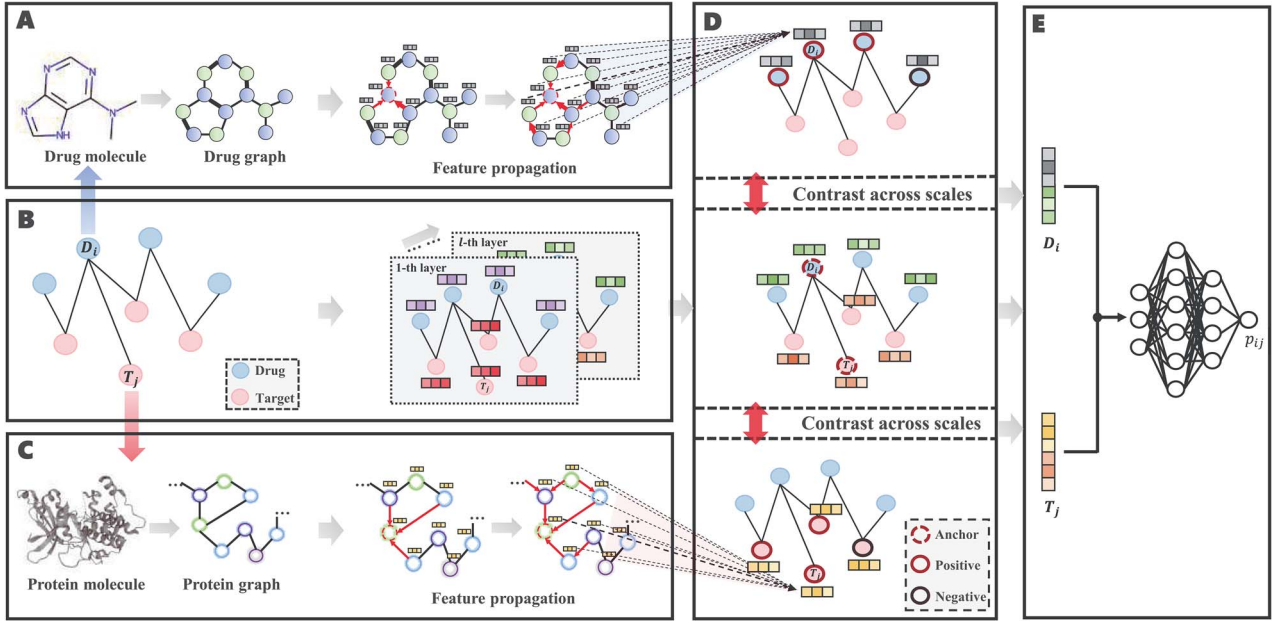


Figure 1. The architecture of CSCo-DTA and its components are shown. A. The drug molecule is modeled as a drug graph and encoded as embeddings by GCNs. B. Graph representation learning in a drug–target network-scale view. C. The target molecule is modeled as a protein graph and encoded as embeddings by GCNs. D. The cross-scale graph contrastive learning component is used to maximize the mutual information between the above two scales. E. The drug–target affinity prediction component consists of an MLP that takes the learned embeddings as input to predict the affinity score.

notation used in our work. drug–target affinity prediction problem, drug–target bipartite network, drug molecular graph, protein molecular graph and the Laplacian matrix are defined as follows:

- 1) **drug–target binding affinity (DTA) prediction.** Given a set of drugs \mathcal{D} and a set of targets \mathcal{T} , the binding affinity values between drugs and targets can be organized into a matrix $\mathbf{E} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{T}|}$. Here, we aim to find a matrix $\hat{\mathbf{E}} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{T}|}$ to approximate \mathbf{E} to predict the unobserved binding affinity value between the drug and the target.
- 2) **drug–target bipartite network.** The bipartite drug–target affinity network could be defined as $\mathcal{G}_N = (\mathcal{D}, \mathcal{T}, \mathcal{E})$, where \mathcal{D} and \mathcal{T} are the set of drugs and targets, respectively, and $e_{ij} \in \mathcal{E}$ indicates the affinity between $D_i \in \mathcal{D}$ and $T_j \in \mathcal{T}$. The weight of edge e_{ij} represents how tightly D_i binds to T_j . The adjacent matrix $\mathbf{A}_{\mathcal{G}_N} \in \mathbb{R}^{(|\mathcal{D}|+|\mathcal{T}|) \times (|\mathcal{D}|+|\mathcal{T}|)}$ corresponding to the drug–target bipartite network can be defined as

$$\mathbf{A}_{\mathcal{G}_N} = \begin{bmatrix} 0 & \mathbf{E} \\ \mathbf{E}^T & 0 \end{bmatrix} \quad (1)$$

- 3) **Drug molecular graph.** Given a drug $D_i \in \mathcal{D}$, the corresponding drug molecular graph can be formulated as $\mathcal{G}_{D_i} = (\mathcal{M}, \mathcal{A})$, where $\mathcal{M} = \{d_m\}_{m=1}^{|\mathcal{M}|}$ is the set of atoms of the drug, and \mathcal{A} is the set of bonds of the drug, $(d_a, d_b) \in \mathcal{A}$ indicates the existence of a bond between the atom d_a and atom d_b .
- 4) **Protein molecular graph.** Each target protein can be regarded as a graph with amino acid residues as nodes. The molecule graph of the target $T_j \in \mathcal{T}$ is denoted as $\mathcal{G}_{T_j} = (\mathcal{N}, \mathcal{C})$, where \mathcal{N} is a set consisting of amino acids and $(t_a, t_b) \in \mathcal{C}$ indicates that there is a contact between t_a and t_b . It is noteworthy that the protein contact map is used in this paper, which is represented by a symmetric boolean matrix as a simplified representation of protein structure. The protein structure’s relevant characteristics can be obtained by analyzing the protein contact map.

Graph representation learning components for different scales

In this part, we employ graph neural networks to obtain feature representations of drugs and proteins from both network-level and molecule-level perspectives based on the drug–target bipartite network and molecular graph, respectively. Details are shown as follows.

Graph representation learning in a drug–target network-scale view

The theory of network pharmacology models the interactions of drugs and their targets from the perspective of system biology. Also, it proves the importance of interactive information in the drug–target network [41]. Therefore, we perform feature extraction operations on the DTA bipartite network by utilizing the graph convolutional network (GCN). The essence of this operation is to dynamically filter the graph signal in the spectral domain to obtain low-frequency signal as the feature representations of drugs and targets in the network.

Mathematically, suppose $\mathbf{A}_{\mathcal{G}_N}$ and \mathbf{F} are the adjacent matrix and initial features corresponding to the DTA bipartite network, respectively, where $\mathbf{F} = \begin{bmatrix} \mathbf{F}^d \\ \mathbf{F}^t \end{bmatrix}$, \mathbf{F}^d and \mathbf{F}^t represent initial features for drugs and targets, respectively. The feature aggregation operator of GCN is defined as follows:

$$\mathbf{F}^{(l)} = \sigma \left(\mathbf{L} \Theta^{(l)} \sigma \left(\dots \sigma \left(\mathbf{L} \Theta^{(1)} \mathbf{F} \right) \dots \right) \right), \quad (2)$$

where $\sigma(\cdot)$ denotes the nonlinear activation function, such as ReLU [42]. $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A}_{\mathcal{G}_N} \mathbf{D}^{-\frac{1}{2}}$ is the normalized Laplacian matrix, \mathbf{I} is the identity matrix and \mathbf{D} is the diagonal degree matrix defined as $\mathbf{D}_{qq} = \sum_k [\mathbf{A}_{\mathcal{G}_N}]_{qk}$. $\Theta^{(l)}$ represents the trainable parameter matrix of the l -th-layer GCN. $\mathbf{F}^{(l)}$ denotes the embedding matrix at the l -th-layer. Equation (2) describes the iterative process of learning the

embedding of each node by aggregating information from their neighborhood and transforming the aggregated embeddings via multiplying a transformation matrix. Each vertex in the network gradually gains information from its high-order neighbors with the iterative process. Further, the final representation of the drug D_i and the target T_j based on the network scale can be obtained as follows:

$$\mathbf{h}_{D_i}^{mv} = \mathbf{F}^{(l)}[i, :]; \quad (3)$$

$$\mathbf{h}_{T_j}^{nv} = \mathbf{F}^{(l)}[|\mathcal{D}| + j, :], \quad (4)$$

where $\mathbf{F}^{(l)}[i, :]$ represents the i -th row of $\mathbf{F}^{(l)}$, and $\mathbf{h}_{D_i}^{mv}$ and $\mathbf{h}_{T_j}^{nv}$ represent, respectively, the embedding of drug D_i and target T_j based on the molecular scale.

Graph representation learning in a molecule-scale view

In order to extract the molecule-level features, drug and target molecules are modeled as undirected graphs. For a drug molecular graph, nodes and edges in the graph represent atoms and bonds between atoms, respectively. For a protein molecular graph, nodes and edges in the graph represent residues and contacts between residues, respectively. The GCN is applied to encode the topology structure of a molecule-level graph.

More concretely, given a drug D_i , let $\mathcal{G}_{D_i} = (\mathcal{M}, \mathcal{A})$ be the corresponding molecular graph (see the Preliminaries section for details). For the atom of a drug, the initial feature consists of several properties, including element type, bonding number, the count of attached hydrogen atoms, valence and whether the atom is an aromatic ring. For detailed information on the initial characteristics of nodes and edges in drug molecules, please refer to [Supplementary Table 1](#). Based on initial atom features and molecular graph structure, we employ GCNs to learn an embedding vector for each atom and obtain the representation of the drug molecule through the readout function. Following the strategy of neighborhood aggregation, GCN produces the feature representation for each atom via aggregating embeddings of its neighbor atoms and itself iteratively. Formally, the feature aggregation in a single-layer GCN for the atom u in the drug molecule is defined as

$$\mathbf{x}_u^{(l+1)} = \sigma \left(\sum_{v \in N_u^D} \frac{1}{c_u c_v} \boldsymbol{\Theta}_D^{(l)} \mathbf{x}_v^{(l)} \right), \quad (5)$$

where $\mathbf{x}_u^{(l)} \in \mathbb{R}^{f_D}$ is the embedding of the atom u in the l -th layer and f_D represents the dimension of the atom feature. $\boldsymbol{\Theta}_D^{(l)}$ is the parameter matrix of the l -th layer GCN. N_u^D is the neighborhood set of the atom u containing itself. $c_u = |N_u^D|$ is a normalization constant.

Given a target T_j , let $\mathcal{G}_{T_j} = (\mathcal{N}, \mathcal{C})$ be the corresponding molecular graph (see the Preliminaries section for details). For the residue of a target, the initial feature consists of position-specific score, residue type, whether the residue is aromatic, residue weight and so on. For detailed information on the initial characteristics of nodes and edges in protein target molecules, please refer to [Supplementary Table 2](#). We also define the feature aggregating operation for an amino acid residue u in the target molecule as follows:

$$\mathbf{y}_u^{(l+1)} = \sigma \left(\sum_{v \in N_u^T} \frac{1}{c_u c_v} \boldsymbol{\Theta}_T^{(l)} \mathbf{y}_v^{(l)} \right), \quad (6)$$

where $\mathbf{y}_u^{(l)} \in \mathbb{R}^{f_T}$ is the representation of the residue u in the l -th layer and f_T represents the dimension of residue feature. $\boldsymbol{\Theta}_T^{(l)}$ represents the parameter matrix of the l -th layer GCN. N_u^T is the neighborhood set of the residue u containing itself. $c_u = |N_u^T|$ is a normalization constant.

To sum up, the representation of each atom (amino acid residue) can be learned by the molecule-scale GCN, and the corresponding feature of the drug (protein) molecular can be further represented by a readout function as follows:

$$\mathbf{h}_{D_i}^{mv} = \Phi \left(\left\{ \mathbf{x}_m^{(l+1)} \mid m \in \mathcal{M} \right\} \right); \quad (7)$$

$$\mathbf{h}_{T_j}^{nv} = \Phi \left(\left\{ \mathbf{y}_n^{(l+1)} \mid n \in \mathcal{N} \right\} \right), \quad (8)$$

where $\Phi()$ is the differentiable readout function, which is an add operation. $\mathbf{h}_{D_i}^{mv}$ and $\mathbf{h}_{T_j}^{nv}$ represent the embedding of drug D_i and target T_j based on the molecular scale. Through this molecule-level feature extraction, we can gain the feature representation of drugs and proteins from a microscopic perspective.

Cross-scale graph contrastive learning module

The molecule scale-based embeddings and the network scale-based embeddings represent drug and target features from different views, which may be related. In order to enhance feature representations by combining the molecule and network scales, we propose a cross-scale graph contrastive learning method to model embeddings based on different scales.

We first introduce the contrastive learning scheme for drugs. The scheme for targets is similar. After obtaining the drug representations of different scales by aforementioned graph neural network-based method, we first map them into the same feature space with an MLP, as shown in the following equation:

$$\tilde{\mathbf{h}}_{D_i}^{mv} = \text{MLP}(\mathbf{h}_{D_i}^{mv}); \quad (9)$$

$$\tilde{\mathbf{h}}_{D_i}^{nv} = \text{MLP}(\mathbf{h}_{D_i}^{nv}), \quad (10)$$

where $\text{MLP}()$ is a two-layer feedforward neural network. It is noted that $\text{MLP}()$ is shared by embeddings of different scales.

Afterwards, as shown in [Figure 1D](#), the proposed model utilizes a contrastive learning scheme to optimize the representations. Given a drug as an anchor, its representation should be consistent with its positive samples and different from its negative samples across different scales. One of the key challenges of graph contrastive learning across scales is to define the positive and negative samples of a given anchor for providing self-supervised signals. In general, given an embedding of a drug $D_i \in \mathcal{D}$ learned based on the network scale, its embedding learned from the molecular scale can be defined as the positive sample. In this study, to generate positive samples, we propose a novel sampling strategy, which includes molecular feature-based sampling and network structure-based sampling. For molecular feature-based sampling, the idea is that nodes with similar molecular structures are considered highly related and can be used as positive sample pairs. Specifically, the PubChem structure clustering tool [43] is used to calculate the molecular level drug-drug similarities, saved as $\mathbf{S}^d \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}|}$, and the Smith-Waterman algorithm [44] is used to calculate the protein sequence similarities, saved as $\mathbf{S}^t \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$. Both drug-drug similarities and protein sequence similarities are molecular-level information. For network structure-based sampling, the idea is that nodes that share common neighbors in the network are related. We propose a meta-path-based method to

calculate the similarity between drugs. The similarities between drugs are defined as the number of meta-path, which is *drug–target–drug*, between drugs in the drug–target network. The drug similarities are normalized and saved as $\mathbf{Q}^d \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}|}$. Similarly, we obtained the target similarities, saved as $\mathbf{Q}^t \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$. Following aforementioned ways, we selected positive samples for each drug and each target according to molecule-level and network-level views. Specifically, to select positive samples for D_i , we summed \mathbf{S}^d and \mathbf{Q}^d , then rank the samples with the i -th row in descending order. The top K samples are selected as positive samples and the others are considered as negative samples.

The embeddings of positive sample pair should be similar. The embeddings of negative sample pair should be different. Therefore, inspired by InfoNCE [45], the contrastive loss function of network scale is defined as follows:

$$\mathcal{L}_{D_i}^{drug_{nv}} = -\log \frac{\sum_{D_u \in \mathbb{P}_{D_i}} \exp\left(s\left(\tilde{\mathbf{h}}_{D_i}^{nv}, \tilde{\mathbf{h}}_{D_u}^{nv}\right) / \tau\right)}{\sum_{D_k \in (\mathbb{P}_{D_i} \cup \mathbb{N}_{D_i})} \exp\left(s\left(\tilde{\mathbf{h}}_{D_i}^{nv}, \tilde{\mathbf{h}}_{D_k}^{nv}\right) / \tau\right)}, \quad (11)$$

where $s(\mathbf{h}_a, \mathbf{h}_b)$ represents the cosine similarity between \mathbf{h}_a and \mathbf{h}_b . τ is a temperature parameter that used to adjust the similarity distribution of embeddings to better distinguish between positive and negative samples, and is typically set empirically. \mathbb{P}_{D_i} and \mathbb{N}_{D_i} denote the positive and negative sample sets of D_i , respectively.

Similarly, the contrastive loss function of molecule scale is defined as follows:

$$\mathcal{L}_{D_i}^{drug_{mv}} = -\log \frac{\sum_{D_u \in \mathbb{P}_{D_i}} \exp\left(s\left(\tilde{\mathbf{h}}_{D_i}^{mv}, \tilde{\mathbf{h}}_{D_u}^{mv}\right) / \tau\right)}{\sum_{D_k \in (\mathbb{P}_{D_i} \cup \mathbb{N}_{D_i})} \exp\left(s\left(\tilde{\mathbf{h}}_{D_i}^{mv}, \tilde{\mathbf{h}}_{D_k}^{mv}\right) / \tau\right)} \quad (12)$$

In similar way, we can also define $\mathcal{L}_j^{target_{nv}}$ and $\mathcal{L}_j^{target_{mv}}$ for the target T_j as follows:

$$\mathcal{L}_{T_j}^{target_{nv}} = -\log \frac{\sum_{T_u \in \mathbb{P}_{T_j}} \exp\left(s\left(\tilde{\mathbf{h}}_{T_j}^{nv}, \tilde{\mathbf{h}}_{T_u}^{nv}\right) / \tau\right)}{\sum_{T_k \in (\mathbb{P}_{T_j} \cup \mathbb{N}_{T_j})} \exp\left(s\left(\tilde{\mathbf{h}}_{T_j}^{nv}, \tilde{\mathbf{h}}_{T_k}^{nv}\right) / \tau\right)} \quad (13)$$

$$\mathcal{L}_{T_j}^{target_{mv}} = -\log \frac{\sum_{T_u \in \mathbb{P}_{T_j}} \exp\left(s\left(\tilde{\mathbf{h}}_{T_j}^{mv}, \tilde{\mathbf{h}}_{T_u}^{mv}\right) / \tau\right)}{\sum_{T_k \in (\mathbb{P}_{T_j} \cup \mathbb{N}_{T_j})} \exp\left(s\left(\tilde{\mathbf{h}}_{T_j}^{mv}, \tilde{\mathbf{h}}_{T_k}^{mv}\right) / \tau\right)} \quad (14)$$

In general, the part of the contrastive loss function is given as follows:

$$\mathcal{L}^{drug} = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \left[\lambda \cdot L_{D_i}^{drug_{nv}} + (1 - \lambda) \cdot L_{D_i}^{drug_{mv}} \right] \quad (15)$$

$$\mathcal{L}^{target} = \frac{1}{|\mathcal{T}|} \sum_{j=1}^{|\mathcal{T}|} \left[\lambda \cdot L_{T_j}^{target_{nv}} + (1 - \lambda) \cdot L_{T_j}^{target_{mv}} \right], \quad (16)$$

where λ is a hyperparameter to balance the effect of both scales.

Objective with multi-task training strategy

We introduce a supervised-based loss by exploiting the original paired DTA data during training, in addition to the cross-scale contrastive-based loss. This approach allows us to fully utilize the existing labeled DTA data and maximize its potential.

In our approach, we learned the representations of drugs and targets from two different scales and projected them to

the embedding space using an MLP. This results in two sets of embeddings for drugs and targets: $\tilde{\mathbf{h}}_{D_i}^{mv}, \tilde{\mathbf{h}}_{D_i}^{nv} \in \mathbb{R}^{|\mathcal{D}| \times f}$, and $\tilde{\mathbf{h}}_{T_j}^{mv}, \tilde{\mathbf{h}}_{T_j}^{nv} \in \mathbb{R}^{|\mathcal{T}| \times f}$. To obtain the final representations for drugs and targets, we concatenated the mapped embeddings of both views, as shown below:

$$\mathbf{h}_{D_i} = \left[\tilde{\mathbf{h}}_{D_i}^{mv} \parallel \tilde{\mathbf{h}}_{D_i}^{nv} \right]; \quad (17)$$

$$\mathbf{h}_{T_j} = \left[\tilde{\mathbf{h}}_{T_j}^{mv} \parallel \tilde{\mathbf{h}}_{T_j}^{nv} \right], \quad (18)$$

where $[\parallel]$ denotes the concatenation operation. We used the concatenated embeddings $\mathbf{h}_{D_i} \in \mathbb{R}^{2f}$ and $\mathbf{h}_{T_j} \in \mathbb{R}^{2f}$ to predict the binding affinity score. The mean squared error (MSE) loss can be represented as follows:

$$\mathcal{L}^{MSE} = \sum_{(i,j) \in \Omega} \frac{1}{|\Omega|} \left\| e_{ij} - \text{MLP}\left(\left[\mathbf{h}_{D_i} \parallel \mathbf{h}_{T_j}\right]\right) \right\|_F^2, \quad (19)$$

where $\text{MLP}()$ is a feedforward neural network and Ω is used to denote the entries in the training set. e_{ij} represents the experimentally verified binding affinity score between D_i and T_j .

To summarize, the mathematical formulation for multi-task training in DTA prediction is as follows:

$$\mathcal{L} = \mathcal{L}^{MSE} + \frac{1}{2} \mathcal{L}^{drug} + \frac{1}{2} \mathcal{L}^{target} \quad (20)$$

By incorporating the cross-scale contrastive training objective into the DTA prediction task, we jointly optimized both objectives to improve the representations of drugs and targets during training.

RESULTS AND DISCUSSION

Data preparation

To comprehensively evaluate the performance of our proposed method, CSCo-DTA, we use two public drug–target affinity benchmark datasets: the Davis dataset [46] and the KIBA dataset [47]. The Davis dataset takes the negative logarithm of dissociation constants (pKd) as affinity values and includes 30 056 values between 68 drugs and 442 targets. The affinity scores range from 5.0 to 10.8 and indicate the binding strength of drug–target pairs. The KIBA dataset comprises 52 498 compounds and 467 kinase targets, and integrates the information of K_d , K_i and IC_{50} as drug–target affinity scores. The dataset was preprocessed following the method described in [30], resulting in 118 254 binding affinity scores between 2111 drugs and 229 targets, ranging from 0.0 to 17.2. In addition, each drug molecule was processed by RDKit [48] and target molecule was processed by PconsC4 [38, 49] to generate compound graph and protein contact map in the molecule-scale view (as described in Methods section). We compared our proposed method with other methods on these two datasets. The datasets were randomly divided into six subsets of equal size, with five subsets used for training, and one subset used as the test set [31, 34–36, 38]. To comprehensively evaluate the performance of the models, the MSE and r_m^2 were used to assess the predictive performance of all methods.

The performance evaluation of CSCo-DTA and baseline methods

We compared the proposed CSCo-DTA with the existing baseline methods, including DeepDTA [31], GraphDTA [36], DGraphDTA [38], MATT_DTI [34] and FusionDTA [35]. DeepDTA utilized CNNs

to extract features from drug and target sequences, which were then used to predict their binding affinities. GraphDTA utilized graph neural networks to embed drug molecular graph and the target sequence embeddings were extracted through CNN. DGraphDTA utilized protein contact maps to represent the interaction between residue pairs in protein molecules, and employed graph neural networks to extract features from both protein contact maps and drug molecule graphs. MATT_DTI used 1D CNNs to learn the sequence representations of drugs and targets, and multiple attention blocks to establish the relationships between them. FusionDTA employed knowledge distillation and a novel multi-head linear attention mechanism, which aggregated global information using attention weights to make DTA predictions. In addition, we used CSCat-DTA for comparison, which directly concatenated the embeddings learned from network-scale and molecule-scale features via GCNs for fusion, and then used an MLP for prediction. The comparison methods were re-implemented based on the descriptions in the original publications. The experimental parameters for all comparison methods were set according to the model parameters given in the paper of each method.

The main hyperparameters of CSCo-DTA are as follows: (i) the temperature parameter τ in the contrastive loss function, (ii) the biased item λ to balance the effect of two scales in the contrastive loss function, (iii) the number of positive samples K for each node in contrastive learning scheme, (iv) the number of GCN layers at the network-scale and (v) the number of GCN layers at the molecule-scale. Specifically, while keeping other parameters fixed, we systematically varied one parameter at a time to assess its influence on MSE to evaluate the effect of these parameters on the performance of CSCo-DTA. The effects of different parameter choices on model performance are presented in [Supplementary Figure 1](#). We used $\tau = 0.8$, $\lambda = 0.5$, $K = 3$, two layers of GCN in the network-scale, three layers of GCN in the molecule-scale and a learning rate of $2e-4$ as experimental settings for CSCo-DTA.

The performances of CSCo-DTA and other methods on the Davis dataset and KIBA dataset are shown in [Table 1](#). The results indicate that CSCo-DTA achieves the best performance compared with the other baseline methods in most cases. Specifically, the results show significant improvement in CSCo-DTA and DGraphDTA compared with DeepDTA, which demonstrates the advantages of modeling drugs and targets as graphs to leverage their molecule structural information compared with modeling them as sequences. Moreover, CSCo-DTA outperforms DGraphDTA on both datasets, achieving 24.89% improvement in MSE and 12.95% improvement in r_m^2 on the Davis dataset and 3.05% improvement in MSE and 1.38% improvement in r_m^2 on the KIBA dataset. These results demonstrate that integrating molecule-scale and network-scale information can enhance drug–target affinity prediction. Compared with FusionDTA, the performance of CSCo-DTA achieves an average relative improvement of 20.19% in MSE and 3.60% in r_m^2 on the Davis dataset. The comparison between CSCo-DTA and CSCat-DTA indicates that the cross-scale graph contrastive learning component effectively improves DTA prediction performance in comparison with directly concatenating molecular scale and network scale features for prediction.

The predicted affinity scores are correlated with the experimental affinity scores

To further evaluate the performance of our proposed drug–target affinity prediction model, we conducted experiments to compare the predicted values with the measured values on two datasets.



Figure 2. Scatter plot showing the correlation between CSCo-DTA’s predicted scores and measured binding affinities on the Davis dataset. The x-axis represents scores predicted by the CSCo-DTA model, and the y-axis represents true measured values.

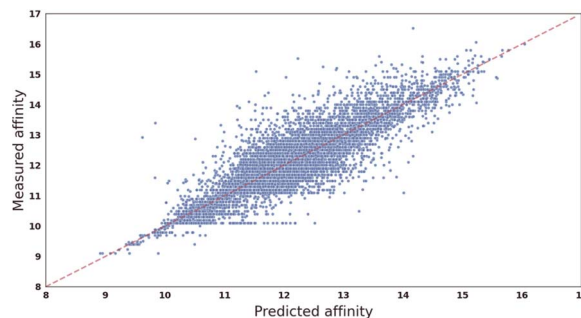


Figure 3. Scatter plot showing the correlation between CSCo-DTA’s predicted scores and measured binding affinities on the KIBA dataset. The x-axis represents scores predicted by the CSCo-DTA model, and the y-axis represents true measured values.

We visually presented the results in the form of scatterplots, as shown in [Figure 2](#) and [3](#), where the x-axis represents the predicted values and the y-axis represents the true measured values. The red dashed line represents the line of perfect agreement between the predicted and true values, and closer data points to this line indicate higher prediction accuracy.

Our analysis reveals that for the Davis dataset, the dense region of data points is within the range of [5, 6], as more than half of the measured values (24 578/30 056) fall within this interval. Similarly, for the KIBA dataset, the dense region is within the range of [10, 14]. The results demonstrate that the CSCo-DTA model predicted the affinity scores very close to the ground-truth values, as evidenced by the data points from both datasets being distributed around the red dashed line. In summary, the experimental results support the high accuracy and reliability of our proposed model in predicting drug–target affinity scores, and highlight its potential in facilitating drug discovery and development efforts.

Multi-scale integration and contrastive learning enable CSCo-DTA to benefit

In this section, we conducted ablation studies to investigate the contribution of each component in our proposed model. Specifically, we examined the effectiveness of different level features and the cross-scale graph contrastive module by removing each component one by one and examining prediction performance. We conducted experiments on the Davis dataset, and the prediction performance results were shown in [Figure 4](#) and [Figure 5](#), which demonstrate the contribution of each component to the overall performance of the proposed model.

The results in [Figure 4](#) and [Figure 5](#) indicate that using only molecule-scale feature (i.e. CSCo-DTA^{−NetFeature}) or network-scale

Table 1: Comparison results of MSE and r_m^2 values for CSCo-DTA and baseline methods, proving our proposed method has improved performance compared with baseline methods on the Davis dataset and KIBA dataset. Lower MSE values indicate better performance. Higher r_m^2 values indicate better performance

Methods	Davis		KIBA	
	MSE ↓	r_m^2 ↑	MSE ↓	r_m^2 ↑
DeepDTA[31]	0.244	0.664	0.180	0.721
GraphDTA[36]	0.242	0.679	0.154	0.736
DGraphDTA[38]	0.221	0.687	0.131	0.797
MATT_DTI[34]	0.228	0.693	0.134	0.786
FusionDTA[35]	0.208	0.749	0.130	0.791
CSCat-DTA(ours)	0.178	0.767	0.130	0.797
CSCo-DTA(ours)	0.166	0.776	0.127	0.808

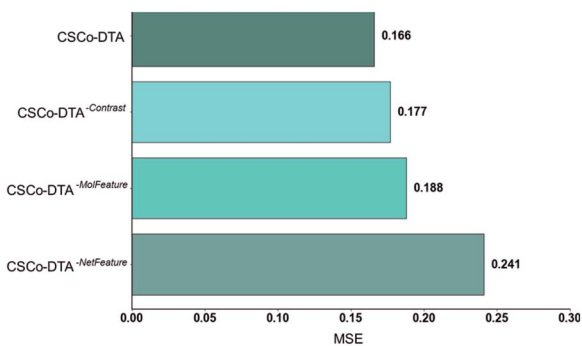


Figure 4. The MSE performance comparison of CSCo-DTA and its three variants on Davis dataset, proving the importance of features at different scales and the fusion of multi-scale features.

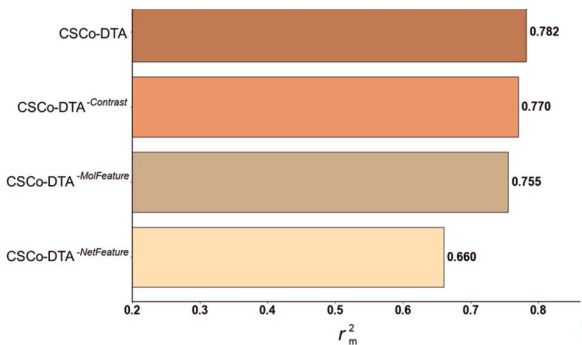


Figure 5. r_m^2 performance comparison of CSCo-DTA and its three variants on Davis dataset, proving the importance of features at different scales and the fusion of multi-scale features.

feature (i.e. CSCo-DTA^{-MolFeature}) leads to decreased performance compared with the original model (CSCo-DTA). Specifically, the MSE values for CSCo-DTA^{-NetFeature} and CSCo-DTA^{-MolFeature} are 0.241 and 0.188, respectively, while the MSE value for CSCo-DTA is 0.166. This finding highlights the importance of both scale features in the performance of model, and underscores the necessity of integrating them to achieve optimal results. In addition, the MSE and r_m^2 of the model without cross-scale graph contrastive component (i.e. CSCo-DTA^{-Contrast}) are reduced to 0.177 and 0.770, respectively. The results indicate that the cross-scale graph contrastive component plays a crucial role in improving the performance of the model. Thus, it can be concluded that each component of our proposed model contributes to the prediction performance of DTA.

Furthermore, we conducted experiments to assess the impact of input features on the performance, and the findings are presented in [Supplementary Table 3](#). The results indicate that both node features and edge weights at the molecular and network scales are crucial to the performance. We further assessed the impact of drug–target network topology on the model’s performance through random edge removal and random edge swapping ([Supplementary Tables 5 and 6](#)).

Case study: predicting the new potential targets of Erlotinib

In this study, we applied the proposed CSCo-DTA model to predict potential targets for drug repositioning and performed molecular docking to validate the predicted targets. Specifically, we predicted the potential targets for Erlotinib, a tyrosine kinase receptor inhibitor used for the treatment of advanced or metastatic pancreatic cancer or non-small cell lung cancer [50–52]. Erlotinib was known to bind to the target receptor EGFR and inhibit its activity [53]. We obtained the potential targets for Erlotinib by ranking protein targets according to the prediction scores from CSCo-DTA, and found that 9 out of the top 10 targets with the highest predicted scores ([Supplementary Table 4](#)) had predicted value ranks consistent with their experimental activity ranks. To further validate our results, we performed molecular docking via LibDock [54] for the remaining target, IRAK4, which yielded a docking score of 118.115 compared with 122.603 for EGFR, suggesting a potential interaction between the drug and IRAK4. The docking results were visualized in [Figure 6](#), which demonstrated that IRAK4 formed intermolecular forces with the amino acid residues of the targets. Furthermore, we performed the same experiment with the other method GraphDTA. The result show that IRAK4 is not included in the top 10 predicted targets for Erlotinib. This also indicates the effectiveness of our method on the drug target prediction. In conclusion, our study successfully predict potential protein targets for drugs using a computational model and validate the predicted target protein by molecular docking. These findings provide valuable insights into the mechanism of action of Erlotinib and its potential applications in the treatment of other diseases.

CONCLUSION

Predicting drug–target affinity is crucial for drug discovery and repurposing. In this study, we propose the CSCo-DTA approach to predict the binding affinity scores of drugs with targets. The proposed approach models the interactions between drugs and targets by learning representations from the drug–target affinity network. In addition, CSCo-DTA utilizes a GCN to extract

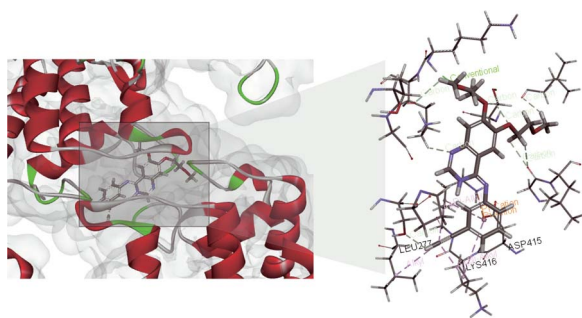


Figure 6. A case of the molecular docking visualization for the interaction between Erlotinib and IRAK4. The left part draws the DTIs. The right part further shows the visualization of the drug and the target pocket.

molecule-level features of drugs or targets, and employs a contrastive learning framework to capture the potential relationship between the network-level and molecule-level views and make the learned representation more informative and integrated. Finally, we optimize the model using an objective with a multi-task training strategy. The evaluation results demonstrate that CSCo-DTA outperforms other methods in terms of MSE and r_m^2 and different components of CSCo-DTA contributes to the performance improvement. We also apply CSCo-DTA to predict potential targets of Erlotinib. The results indicate that the CSCo-DTA may lead to the potential expansion of the therapeutic application of drugs.

Key Points

- We present a novel cross-scale contrastive learning-based method (CSCo-DTA) for predicting drug–target binding affinities.
- CSCo-DTA utilizes graph convolutional network encoders to extract molecule-scale and network-scale features of drugs and targets. A contrastive learning framework is employed to maximize the mutual information between the features of two scales and explore their potential relationship.
- The experimental results show that the CSCo-DTA outperforms state-of-the-art methods on both Davis dataset and KIBA dataset.

SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/bib>.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for valuable suggestions. This paper was supported by the National Natural Science Foundation of China (62072376), Guangdong Basic and Applied Basic Research Foundation (2022A1515010144), Innovation Capability Support Program of Shaanxi (2022KJXX-75), the Fundamental Research Funds for the Central Universities (D5000230056).

CODE AND DATA AVAILABILITY

The implemented code and experimental dataset are available online at <https://github.com/23AIBox/23AIBox-CSCo-DTA>.

REFERENCES

1. DiMasi JA. Assessing pharmaceutical research and development costs. *JAMA Intern Med* 2018;**178**(4):587.
2. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* 2020;**323**(9):844–53.
3. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016;**47**:20–33.
4. Mullard A. New drugs cost US\$2.6 billion to develop. *Nat Rev Drug Disc* 2014;**13**:877.
5. Duarte Y, Márquez-Miranda V, Miossec MJ, González-Nilo F. Integration of target discovery, drug discovery and drug delivery: a review on computational strategies. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* 2019;**11**(4):e1554.
6. Agamah FE, Mazandu GK, Hassan R, et al. Computational/in silico methods in drug target and lead prediction. *Brief Bioinform* 2020; **21**(5):1663–75.
7. Blundell TL, Jhoti H, Abell C. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* 2002;**1**(1):45–54.
8. Bruno A, Costantino G, Sartori L, Radi M. The in silico drug discovery toolbox: applications in lead discovery and optimization. *Curr Med Chem* 2019;**26**(21):3838–73.
9. Honkala A, Malhotra SV, Kummur S, Junttila MR. Harnessing the predictive power of preclinical models for oncology drug development. *Nat Rev Drug Discov* 2022;**21**(2):99–114.
10. Rubin EH, Gilliland DG. Drug development and clinical trials—the path to an approved cancer drug. *Nat Rev Clin Oncol* 2012;**9**(4):215–22.
11. Arrowsmith J. Trial watch: phase II failures: 2008–2010. *Nat Rev Drug Discov* 2011;**10**(5):328–9.
12. Du X, Li Y, Xia YL, et al. Insights into protein–ligand interactions: mechanisms, models, and methods. *Int J Mol Sci* 2016;**17**(2):144.
13. Campbell SJ, Gold ND, Jackson RM, Westhead DR. Ligand binding: functional site location, similarity and docking. *Curr Opin Struct Biol* 2003;**13**(3):389–95.
14. Dhakal A, McKay C, Tanner JJ, Cheng J. Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. *Brief Bioinform* 2022;**23**(1):bbab476.
15. Ma W, Yang L, He L. Overview of the detection methods for equilibrium dissociation constant KD of drug–receptor interaction. *J Pharm Anal* 2018;**8**(3):147–52.
16. Alves NJ, Stimple SD, Handlogten MW, et al. Small-molecule-based affinity chromatography method for antibody purification via nucleotide binding site targeting. *Anal Chem* 2012;**84**(18):7721–8.
17. Vivo MD, Masetti M, Bottegoni G, et al. Role of molecular dynamics and related methods in drug discovery. *J Med Chem* 2016;**59**(9):4035–61.
18. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;**3**(11):935–49.
19. Gorgulla C, Boeszoermyeni A, Wang ZF, et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 2020;**580**(7805):663–8.
20. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**(7873):583–9.
21. Peng J, Wang Y, Guan J, et al. An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. *Brief Bioinform* 2021;**22**(5):bbaa430.

22. Luo Y, Zhao X, Zhou J, et al. A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):573.
23. Wan F, Hong L, Xiao A, et al. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* 2019;**35**(1): 104–11.
24. An Q, Yu L. A heterogeneous network embedding framework for predicting similarity-based drug–target interactions. *Brief Bioinform* 2021;**22**(6):bbab275.
25. Ye Q, Hsieh CY, Yang Z, et al. A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun* 2021;**12**(1):6775.
26. Li Y, Qiao G, Wang K, Wang G. Drug–target interaction prediction via multi-channel graph neural networks. *Brief Bioinform* 2022;**23**(1):bbab346.
27. Zhao T, Hu Y, Valsdottir LR, et al. Identifying drug–target interactions based on graph convolutional network and deep neural network. *Brief Bioinform* 2021;**22**(2):2141–50.
28. Li H, Leung KS, Wong MH, Ballester P. Low-quality structural and interaction data improves binding affinity prediction via random Forest. *Molecules* 2015;**20**(6):10947–62.
29. Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug–target interaction predictions. *Brief Bioinform* 2015;**16**(2): 325–37.
30. He T, Heidemeyer M, Ban F, et al. SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *J Chem* 2017;**9**(1):24.
31. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;**34**(17): i821–9.
32. Lin X, Zhao K, Xiao T, et al. DeepGS: Deep Representation Learning of Graphs and Sequences for Drug–Target Prediction. *The 24th European Conference on Artificial Intelligence (ECAI)*. Santiago de Compostela, Spain, 2020, 1301–8.
33. Li S, Wan F, Shu H, et al. MONN: a multi-objective neural network for predicting compound–protein interactions and affinities. *Cell Systems* 2020;**10**(4):308–322.e11.
34. Zeng Y, Chen X, Luo Y, et al. Deep drug–target binding affinity prediction with multiple attention blocks. *Brief Bioinform* 2021;**22**(5):bbab117.
35. Yuan W, Chen G, Chen C. FusionDTA: attention-based feature polymerizer and knowledge distillation for drug–target binding affinity prediction. *Brief Bioinform* 2022;**23**(1):bbab506.
36. Nguyen T, Le H, Quinn TP, et al. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021;**37**(8):1140–7.
37. Zhang S, Jiang M, Wang S, et al. SAG-DTA: prediction of drug–target affinity using self-attention graph network. *Int J Mol Sci* 2021;**22**(16):8993.
38. Jiang M, Li Z, Zhang S, et al. Drug–target affinity prediction using graph neural network and contact maps. *RSC Adv* 2020;**10**(35): 20701–12.
39. Sun M, Zhao S, Gilvary C, et al. Graph convolutional networks for computational drug development and discovery. *Brief Bioinform* 2020;**21**(3):919–35.
40. Zeng X, Zhu S, Lu W, et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chem Sci* 2020;**11**(7):1775–97.
41. Nogales C, Mamdouh ZM, List M, et al. Network pharmacology: curing causal mechanisms instead of treating symptoms. *Trends Pharmacol Sci* 2022;**43**(2):136–50.
42. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. *J Mach Learn Res* 2011;**15**:315–23.
43. Bolton EE, Wang Y, Thiessen PA, et al. PubChem: integrated platform of small molecules and biological activities. *Annu Rep Comput Chem* 2008;**4**:217–41.
44. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;**147**(1):195–7.
45. Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748. 2018.
46. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;**29**(11):1046–51.
47. Tang J, Szwarzda A, Shakyawar S, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model* 2014;**54**(3):735–43.
48. Lovrić M, Molero JM, Kern R. PySpark and RDKit: moving towards big data in cheminformatics. *Mol Inform* 2019;**38**(6):e1800082.
49. Michel M, Hurtado DM, Elofsson A. PconsC4: fast, accurate and hassle-free contact predictions. *Bioinformatics* 2019;**35**(15): 2677–9.
50. Yang Z, Hackshaw A, Feng Q, et al. Comparison of gefitinib, erlotinib and afatinib in non-small cell lung cancer: a meta-analysis. *Int J Cancer* 2017;**140**(12):2805–19.
51. Cai H, Wang R, Guo X, et al. Combining gemcitabine-loaded macrophage-like nanoparticles and Erlotinib for pancreatic cancer therapy. *Mol Pharm* 2021;**18**(7):2495–506.
52. Mountzios G, Syrigos KN. A benefit-risk assessment of erlotinib in non-small-cell lung cancer and pancreatic cancer. *Drug Saf* 2011;**34**(3):175–86.
53. Bareschino MA, Schettino C, Troiani T, et al. Erlotinib in cancer treatment. *Ann Oncol* 2007;**18**:vi35–41.
54. Rao SN, Head MS, Kulkarni A, LaLonde JM. Validation studies of the site-directed docking program LibDock. *J Chem Inf Model* 2007;**47**(6):2159–71.