

Systems biology

IDDkin: network-based influence deep diffusion model for enhancing prediction of kinase inhibitors

Cong Shen ¹, Jiawei Luo^{1,*}, Wenjue Ouyang¹, Pingjian Ding² and Xiangtao Chen¹

¹College of Computer Science and Electronic Engineering, Hunan University, Changsha 410083, China and ²Division of Biomedical Informatics, University of South China, Hengyang 421001, China

*To whom correspondence should be addressed. luojiawei@hnu.edu.cn

Associate Editor: Alfonso Valencia

Received on July 9, 2020; revised on November 9, 2020; editorial decision on December 6, 2020; accepted on December 10, 2020

Abstract

Motivation: Protein kinases have been the focus of drug discovery research for many years because they play a causal role in many human diseases. Understanding the binding profile of kinase inhibitors is a prerequisite for drug discovery, and traditional methods of predicting kinase inhibitors are time-consuming and inefficient. Calculation-based predictive methods provide a relatively low-cost and high-efficiency approach to the rapid development and effective understanding of the binding profile of kinase inhibitors. Particularly, the continuous improvement of network pharmacology methods provides unprecedented opportunities for drug discovery, network-based computational methods could be employed to aggregate the effective information from heterogeneous sources, which have become a new way for predicting the binding profile of kinase inhibitors.

Results: In this study, we proposed a network-based influence deep diffusion model, named IDDkin, for enhancing the prediction of kinase inhibitors. IDDkin uses deep graph convolutional networks, graph attention networks and adaptive weighting methods to diffuse the effective information of heterogeneous networks. The updated kinase and compound representations are used to predict potential compound-kinase pairs. The experimental results show that the performance of IDDkin is superior to the comparison methods, including the state-of-the-art kinase inhibitor prediction method and the classic model widely used in relationship prediction. In experiments conducted to verify its generalizability and in case studies, the IDDkin model also shows excellent performance. All of these results demonstrate the powerful predictive ability of the IDDkin model in the field of kinase inhibitors.

Availability and implementation: Source code and data can be downloaded from <https://github.com/CS-BIO/IDDkin>.

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The perturbation of protein kinase-mediated cellular signalling pathways causes a number of diseases, including inflammation, cancer and diabetes (Manning, 2002; Noble *et al.*, 2004). Accumulated studies have shown that kinases can be potential targets of drugs for the treatment of various types of human diseases because of their important roles in many cellular activities (Noble *et al.*, 2004). As of June 2020, 61 kinase inhibitors have been approved by the US Food and Drug Administration, and many kinase inhibitors are currently in preclinical and clinical development (Roskoski, 2020). Meanwhile, 518 kinases are included in the human kinome, accounting for ~1.7% of all human genes. However, only a small number of human kinases (~80) have been identified as targets for drugs, and there are many kinase inhibitor drugs that target the same kinases (Fabbro *et al.*, 2015). Approximately 25% of the kinases have completely unknown functions, and nearly half of the

kinases are largely uncharacterized (Bhullar *et al.*, 2018). Thus, a large number of untargeted kinase-related diseases need further research.

Due to the conservation of the ATP binding site, the majority of kinase inhibitors have low selectivity, and this can cause adverse side effects (Metz *et al.*, 2011). Therefore, a detailed study of kinase inhibitor-target interactions is important for understanding their molecular modes of action and provides an opportunity to identify new starting points for other therapeutically interesting kinases (Janssen *et al.*, 2019). Utilizing wet experiments to study the binding profile of kinase inhibitors is still the predominant approach, but even for large-scale pharmaceutical companies, filtering large amounts of compounds in this way is extremely slow and costly (Dickson and Gagnon, 2004; Merget *et al.*, 2017). To accelerate research on the binding profile of kinase inhibitors, it is urgent to develop new methods to compensate for the shortcomings of traditional methods.

Recent advances in computation-based drug discovery methods have enabled the discovery of the potential interactions of thousands of compounds against a range of targets (Merget *et al.*, 2017). Inspired by this, a large number of in silico modelling methods for discovering kinase inhibitory activity have been proposed. Unlike traditional drug design computational methods (including classical QSAR methods and free energy calculation tools), machine learning-based methods, including random forest (RF) (Bora *et al.*, 2016; Cao *et al.*, 2013; Merget *et al.*, 2017), k-nearest neighbours (KNN) (Schurer and Muskal, 2013), naïve Bayesian (NB) (Nijima *et al.*, 2012), deep neural network (DNN) (Li *et al.*, 2019; Manallack *et al.*, 2002) and support vector machine (SVM) (Yabuuchi *et al.*, 2011), have significant advantages in predicting the biological activity of a large number of kinase inhibitors. In general, these methods are trained on separate datasets related to a specific task and they have achieved good results. For example, Cao *et al.* (2013) developed a random forest (RF) method to measure the differentiation of the quantitative binding affinities of kinase-inhibitor pairs. The innovation of Cao's model was that the author proposed a method to represent protein kinases by their amino acid sequences, and the effectiveness of the method was demonstrated through experiments. Nijima *et al.* (2012) proposed a deconvolution approach by constructing dual-component support vector machines (DCSVMs) and dual-component naïve Bayes (DCNB) to dissect kinase profiling data, and this approach enables both the extraction of residue-fragment pairs that are associated with activity and activity prediction of given compounds on a kinome-wide scale. Yabuuchi *et al.* (2011) utilized the machine learning methods of multiple CPIs, which were mainly SVM models, to demonstrate novel lead compounds for the protein kinases and G-protein coupled receptors. Recently, Li *et al.* (2019) presented a multitask deep neural network (MTDNN) method to predict kinase inhibitory activity for large-scale compound data, and the feasibility of this method was demonstrated by various experiments. A prediction model based on the traditional machine learning method accelerates the prediction process of the biological activity of potential kinase inhibitors and makes up for the deficiencies of the traditional wet experimental methods. However, the superiority of machine learning methods depends heavily on the selection of effective features, which may introduce limitations into machine learning methods. Most of these methods take the structural features of inhibitors (compounds) as their input and use machine learning methods as classifiers to output the classification results. The prediction models that only consider the structural features and association relationships of compounds may have limitations in their accuracy and generalization ability for predicting kinase inhibitors. Simultaneously, the emergence of large-scale, biologically heterogeneous networks has provided unprecedented opportunities for many research fields in drug discovery. Thus, it may be feasible to enhance the prediction of kinase inhibitors from the perspective of heterogeneous networks.

Accumulated studies have shown that network-based approaches could accelerate drug discovery and help us quantify the relationships between multiple entities (Barabasi and Oltvai, 2004; Cheng *et al.*, 2019). Meanwhile, the emergence of a large number of novel graph mining methods has made rapid progress in the fields of drug-target prediction (Luo *et al.*, 2017; Nguyen *et al.*, 2019; Shen *et al.*, 2020a,b), drug repositioning (Cheng *et al.*, 2018; Xuan *et al.*, 2019) and drug combinations (Cheng *et al.*, 2019; Ding *et al.*, 2019). In recent years, a series of methods involving graph neural networks (GNNs) has achieved excellent results in effectively fusing network information. Kong and Yu (2020) incorporated the GEDFN architecture to construct a forest graph-embedded deep feedforward network (forgeNet) model for feature graph construction. Pittala and Bailey-Kellogg (2020) proposed a framework that is a unified deep learning-based model for predicting antigen- and antibody-binding interfaces. Long *et al.* (2020) utilized a graph convolutional network-based model to predict potential microbe-drug associations. Tsubaki *et al.* (2019) combined a convolutional neural network (CNN) for proteins and a graph neural network (GNN) to construct a novel CPI prediction approach. Thus, we attempted to predict the biological activity of kinase inhibitors from the

perspective of graph neural networks, which may aggregate more heterogeneous network information and improve the performance of traditional machine learning methods in kinase inhibitor prediction. However, how to utilize the graph neural network method to extract useful knowledge from the kinase inhibitor heterogeneous network is the main challenge of the network-based approach.

To overcome these challenges, we proposed a network-based influence deep diffusion model, named IDDkin, to enhance the prediction of kinase inhibitors. We first constructed a heterogeneous network by integrating the inhibitor (compound) similarity network and the compound-kinase pair network. Then, we used the graph convolution network (GCN) to fuse high-order neighbour information into the compound similarity network. Meanwhile, the graph attention network (GAT) and adaptive weighting were utilized to diffuse the compound-kinase pair information to the kinase nodes and compound nodes. Finally, the process of homogeneous fusion and heterogeneous diffusion was iterated in the heterogeneous networks to obtain deeper, more effective information. The evaluation results demonstrated that IDDkin obtains better results than many state-of-the-art methods in a 5-fold cross-validation. Moreover, the case studies further illuminated IDDkin's strong power for predicting novel kinase inhibitors.

2 Materials and methods

2.1 Framework overview

The overall architecture of IDDkin is illustrated in Figure 1. IDDkin contains a graph convolution network that takes the compound similarity network as its input, a graph attention network that incorporates the effective information of known compound-kinase pairs for kinase representation, and an adaptive weighting model that aggregates the network knowledge of known kinase-compound pairs to update the compounds. These three parts of the IDDkin model synergistically affect the prediction performance of the model. The representation of compounds learned by the graph convolution network provides the initial input at the graph attention network level. The feature incorporation at the graph attention network level and the adaptive weighting level ensures the representation vectors of the compounds and kinases contain valid information for known associations.

2.2 Construction of the heterogeneous network and feature engineering

In this study, the heterogeneous networks consisted of a network of compound similarities and a network of kinase-compound pairs. First, we downloaded compound-target associations from the BindingDB database (Gilson *et al.*, 2016) to construct the compound similarity network and used the Tanimoto coefficient to calculate the similarity of the compounds. Then, we collected the kinase-compound pairs from the Tang set (Tang *et al.*, 2014) and the PKIS set (Elkins *et al.*, 2016; Knapp *et al.*, 2013). Each kinase was divided into the inactive (negative) or active (positive) class through the pIC50 cut-off set at 6.3 (corresponding to 500 nM). Meanwhile, the similarity network was filtered according to the Tang set and PKIS set. Finally, the heterogeneous network based on the Tang set contained 188 kinases, 1351 compounds and 15660 data points. Similarly, the heterogeneous network constructed based on the PKIS set contained 195 kinases, 366 compounds and 2414 data points. The chemical structure (SMILES format) of a compound contains a large amount of physicochemical property information, and it is often used to represent the chemical structure of the compound for drug discovery-related tasks, including molecular generation (Moret *et al.*, 2020; Zang and Wang, 2020), compound structure feature extraction (Shen *et al.*, 2020a,b), and compound similarity calculation (Ding *et al.*, 2020; Luo *et al.*, 2020a,b). Therefore, for the structural features of the compounds, we assembled the chemical structure information (SMILES format) from the Tang set and the PKIS set. RDKit (<http://www.rdkit.org/>) was used to compute the MACCS fingerprints for all of the compounds. An array with a length of 166 bits was obtained to denote each compound.

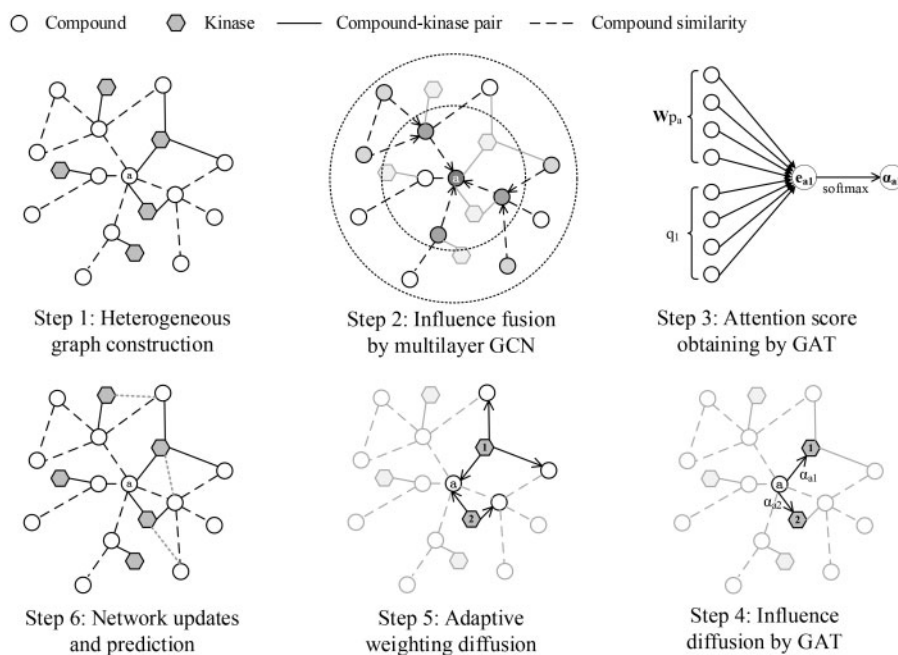


Fig. 1. The overall architecture of IDDkin model. Step 1: Construct a heterogeneous information network by combining compound similarity networks and compound-kinase pair networks. Step 2: Aggregate multi-hop neighbour information in compound similarity network through multi-layer GCN model. Step 3: Calculate the attention score through the compound's representation vector and free embedding of kinase. Step 4: Diffuse the influence from the compound node to the representation of the kinase node by the GAT. Step 5: Diffuse the influence from the kinase node to the representation of the compound node by the adaptive weighting. Step 6: Update all node representations in the network to enhance prediction of kinase inhibitors. The red dashed lines indicate the newly predicted associations

2.3 IDDkin model

2.3.1 Problem formulation

IDDkin predicts potential kinase-compound pairs from a heterogeneous network, which contains compounds, kinases, compound similarities and known kinase-compound pairs. Let $G = (V, E, F)$ denote the undirected heterogeneous network, where V is the set of vertices (compounds and kinases), E is the set of edges (compound similarities and kinase-compound pairs), and F is the set of node embeddings. Thus, there are two sets of entities: compound set A ($A = N_A$) and kinase set B ($B = N_B$). Furthermore, the compound similarity matrix can be represented as $S \in \mathbb{R}^{N_A \times N_A}$. Let $Y \in \mathbb{R}^{N_A \times N_B}$ represent the association matrix of the N_A known inhibitors against a panel of N_B kinase assays, $Y_{ij} = 1$ if compound A_i is tested against kinase B_j at an activity threshold of $pIC_{50} = 6.3$; otherwise, it equals 0. In addition, each compound is associated with real-valued attributes (e.g. compound structure features), represented as p in the compound structure feature matrix $P \in \mathbb{R}^{N_A \times d}$, where $d = 166$ is the dimension of the compound structure feature. To capture the network topology information, let $Q \in \mathbb{R}^{N_B \times L}$ denote the free embedding of the kinases, where L represents the dimension of the free embedding. Given the random initial values of the free embedding, IDDkin updates the stability through optimization operations, which can capture the collaborative latent representation of the compounds and kinases (Wu *et al.*, 2019).

2.3.2 Homogeneous fusion layer

In recent decades, accumulated studies (Bleakley and Yamanishi, 2009; Luo *et al.*, 2017) have found that similarity-based calculation methods have achieved good performance for computational drug-target interaction (DTI) prediction. A common assumption of these methods is the 'guilt-by-association', that is, similar drugs may share similar targets and vice versa. Inspired by these methods for DTI prediction, we applied the assumption based on 'guilt-by-association' to the task of kinase inhibitor prediction. Thus, we have constructed a compound similarity network and look forward to making full use of the effective information in the compound similarity network.

In a similarity network, similarity information between compounds contains a large amount of topological information, which can facilitate the representation learning of specific tasks. In particular, a large amount of effective neighbour information is included in the most similar neighbours, and the low-similarity neighbours have little effect on the model (Zhang *et al.*, 2020). First, we use the K-nearest neighbour of each node to filter its neighbour information as follows:

$$S'_{ij} = \begin{cases} S_{ij}^N & a_j \in N(a_i) \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $N(a_i)$ denotes the K-nearest neighbour of a_i and S^N is row normalized by S . Then, we develop an information fusion method-based graph convolutional network (Kipf and Welling, 2017) to aggregate the neighbour information of each compound as follows:

$$\text{GCN}(S', P) = \sigma(\sim D^{-\frac{1}{2}} \sim S \sim D^{-\frac{1}{2}} P W) \quad (2)$$

where $\sim S = S' + I$ is the similarity matrix with self-connection for the compound similarity network, $W \in \mathbb{R}^{d \times d}$ is the weight matrix, $\sim D$ is the diagonal degree matrix of $\sim S$ and $\sigma(\cdot)$ denotes the activation function. To aggregate the multi-hop neighbour information, a deep graph convolutional network is used in the IDDkin model:

$$P^{l+1} = \text{GCN}^l(S', P^l) \quad (3)$$

$$\text{GCN}^l(S', P^l) = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{S} \tilde{D}^{-\frac{1}{2}} P^l W^l) \quad (4)$$

where l represents the number of layers in the deep graph convolutional network.

The whole process of the deep graph convolutional network corresponds to step 2 in Figure 1. Taking compound a as an example, the dark nodes in the first-order neighbours that need to be aggregated by the graph convolutional network and the nodes with a light colour are those that are removed from the K-nearest neighbour. The same screening method is extended to the deep neighbours of compound a . As the number of layers increases, the noise

information in the aggregated deep neighbours will also increase, and the performance with different depths is illustrated in the section describing the detailed model analysis. Compound features that incorporate deep neighbour information will be used as input for the heterogeneous diffusion layer in the next stage.

2.3.3 Heterogeneous diffusion layer

In IDDkin, the effective information of the kinase nodes mainly comes from network knowledge, especially the association between kinases and compounds. It is worth noting that the inhibitory intensity of a compound for different kinases is different in compound-kinase pairs with a known inhibitory effect. Therefore, a method that can adaptively learn the inhibitory strength of compounds against kinases is needed to solve this problem.

A graph attention network (Veličković et al., 2018) is applied to the interaction graph without edge attributes. In recent years, graph attention networks have also been widely used in bioinformatics (Du et al., 2020) and drug screening (Wang et al., 2020). Thus, we used a graph attention network to adaptively learn the inhibitory strength of the compounds on the kinases in IDDkin. With the learned compound feature vector P_{l+1} as input, free embedding of the kinase is calculated by the neighbour incorporation of the interaction graph as follows:

$$\text{Agg}(P) = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W P_j^l\right) \quad (5)$$

where $\sigma(\cdot)$ denotes the non-linear activation function, and $N(i)$ represents the set of compound nodes that interact with kinase node i . W is the weight matrix, and P_j^l is the j th row of the compound feature matrix after gathering neighbour information through the graph convolutional network. It is worth noting that α_{ij} is the attention score between kinase node i and compound node j , which can be calculated by the following equation:

$$\alpha_{ij} = \frac{\exp(\text{LeakyRelu}(\mathbf{a}(\text{concat}(\mathbf{Q}_i, [\mathbf{W}^P \mathbf{P}_j])))})}{\sum_{n \in N(i)} \exp(\text{LeakyRelu}(\mathbf{a}(\text{concat}(\mathbf{Q}_i, [\mathbf{W}^P \mathbf{P}_n])))})} \quad (6)$$

where $\text{concat}(\cdot)$ is the concatenation operation and \mathbf{a} is a learnable attention weight vector. W^P is the projection matrix of the compound node space. The attention score is calculated as shown in step 3 in Figure 1. Afterward, we connected the compound diffusion presentation $\text{Agg}(P)$ and free embedding Q as follows:

$$Q_f = \tau(\text{concat}(Q, \text{Agg}(P)) \times W_Q) \quad (7)$$

where $\text{concat}(\cdot)$ represents the concatenate operation and $\tau(\cdot)$ is the sigmoid function. $W_Q \in \mathbb{R}^{(d+L) \times L}$ is the weight matrix, and Q_f is the representation vector of the kinase after the diffusion of the influence through the compound nodes. The process of influence diffusion is shown in step 3 in Figure 1.

After the IDDkin model executes the graph convolutional network and graph attention network, the representation vector of the kinase nodes has simultaneously merged the topology information in the compound similarity network with the network knowledge in the compound-kinase interaction network, but the compound representation has not yet integrated the network knowledge in the interaction network. Meanwhile, kinases are also important neighbours of compounds. Thus, we will diffuse the representation of the kinase nodes that combines the information of the similarity network and the interaction network to the compound nodes. We use adaptive weighting to achieve this goal as follows:

$$P'_{ij} = \sigma(\beta_i \times (Q_f)_{:j}) \quad (8)$$

where $\beta \in \mathbb{R}^{N_A \times N_B}$ is the adaptive weight matrix and $\sigma(\cdot)$ is the ReLU function. We connected the homogeneous fusion compound presentation $P^l \in \mathbb{R}^{N_A \times d}$ with the heterogeneous diffusion compound representation $P \in \mathbb{R}^{N_A \times L}$ as follows:

$$P_f = \tau(\text{concat}(P^l, P) \times W_P) \quad (9)$$

where $W_P \in \mathbb{R}^{(d+L) \times d}$ is the weight matrix and $\tau(\cdot)$ denotes the sigmoid function. The matrix $P_f \in \mathbb{R}^{N_A \times d}$ is the compound representation, which aggregates the information of the similarity network and the interaction network. The process of adaptive weighting is shown in step 4 in Figure 1.

2.3.4 Deep IDDkin model

In the IDDkin model, we used a graph convolutional network to incorporate neighbour information into the compound similarity network for the first time and diffused the influence of compounds that aggregate neighbour information through the graph attention network to the kinases for the second time. Finally, we made the updated kinases diffuse influence to the compounds again through adaptive weighting. We tried to combine the deep graph convolutional network from Section 2.3.2 and the graph attention network from Section 2.3.3 to synergistically update the representation of the compounds and kinases. The depth of the deep IDDkin model is mainly reflected in the deep graph convolution network implemented on the compound similarity network. Information fuses in the homogeneous network and diffuses in the heterogeneous network from time to time. Therefore, whether it is a compound or a kinase, its representation after $(l+1)$ th iterations should be a combination of its own representation and its neighbours' representations. The calculation process is expressed as follows:

$$P_f^{l+1} = \tau(\text{concat}(P^{l+1}, \beta \times \tau(\text{concat}(Q, \text{Agg}(P^{l+1})) \times W_Q)) \times W_P) \quad (10)$$

$$Q_f^{l+1} = \tau(\text{concat}(Q, \text{Agg}(P^{l+1})) \times W_Q) \quad (11)$$

$$\text{Agg}(P^{l+1}) = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W(\text{GCN}^l(S', P^l))_j\right) \quad (12)$$

where $\tau(\cdot)$ and $\sigma(\cdot)$ represent the sigmoid function and the ReLU function, respectively. P_f^{l+1} and Q_f^{l+1} denote the $(l+1)$ layer deep presentation of compounds and kinases.

2.3.5 Model training

With the defined presentation matrix of the compounds and kinases in a given heterogeneous network, our idea is that if the two different types of nodes are linked, the prediction score should be close to 1; otherwise, it should be 0. Therefore, there are two challenges: (i) the dimension of the compound's presentation matrix P and the kinase's presentation matrix Q are not uniform; and (ii) how to effectively utilize the known compound-kinase pair information in the model training process. In view of the above challenges, we propose a loss function for model training as follows:

$$\min_{\{Q, G_P, G_Q, \Phi_W, T_x\}} \sum_{\substack{i \in [1, N_A] \\ j \in [1, N_B]}} \left\| Y_{ij} - (P^{l+1})_i \cdot G_P \cdot (G_Q)^T (Q^{l+1})_j^T \right\|_F^2 \quad (13)$$

where $\Phi_W = \{W^l, W, W^P, W_Q, W_P\}$, $T_x = \{\alpha, a, \beta\}$. $G_P \in \mathbb{R}^d \times p$ and $G_Q \in \mathbb{R}^L \times p$ denote the projection matrix of compound node space and the kinase node space, respectively. The above loss function states that after projections of P^{l+1} and Q^{l+1} by G_P and G_Q , respectively, we used the inner product of the two projected vectors to reconstruct the original edge weight Y_{ij} . It is worth noting that such a reconstruction strategy has been used in the previous literature (Luo et al., 2017; Natarajan and Dhillon, 2014; Wan et al., 2019) to solve link prediction problems.

Based on the expression of the loss function, we observed that the prediction of kinase inhibitors is similar to the problem of matrix factorization or completion. However, the main difference from traditional matrix factorization methods (Luo et al., 2017; Natarajan and Dhillon, 2014; Wan et al., 2019) is that the IDDkin

model obtains deeper features of compounds and kinases through fusion and diffusion models. In addition, through the fusion and diffusion steps, IDDkin incorporates the network topology knowledge and compound structure information into feature matrices P^I and Q^I and utilizes these features to guide the next optimization process.

3 Results

3.1 Datasets

Two datasets, Tang *et al.*, (2014) and PKIS (Elkins *et al.*, 2016; Knapp *et al.*, 2013), were used to generate the activity prediction models. The Tang dataset (Tang *et al.*, 2014) was collected from the kinase profiling datasets of Davis *et al.* (2011), Anastassiadis *et al.* (2011) and Metz *et al.* (2011). After initial filtering, the Tang dataset contained 1351 compounds, 188 kinases and a total of 119445 data points in the form of pIC_{50} values (>50% coverage). By setting the pIC_{50} cut-off value to 6.3 (corresponding to 500 nM), the kinases were divided into inactive (negative) and active (positive) classes. Each kinase was divided into the inactive (negative) or active (positive) class through the pIC_{50} cut-off set to 6.3 (corresponding to 500 nM). Therefore, the data points used to construct a heterogeneous network in the Tang dataset changed from 119445 to 15660 (Table 1). PKIS is an abbreviation for Published Kinase Inhibitor Set (Elkins *et al.*, 2016; Knapp *et al.*, 2013), containing 366 compounds, 195 kinases and a total of 71 369 data points in the form of pIC_{50} values (100% coverage). Similar to the processing method for the Tang dataset, the kinases were classified into inactive (negative) and active (positive) classes with the pIC_{50} cut-off value set to 6.3 (corresponding to 500 nM). Similarly, only 2414 data points were used to construct the heterogeneous network for creating the activity prediction models (Table 1).

3.2 Baselines

The proposed IDDkin was compared with the following models: MTDNN (Li *et al.*, 2019), PEPECFP (Avram *et al.*, 2018), KNN (Schurer and Muskal, 2013), Naïve Bayes (Schurer and Muskal, 2013) and Merget's method (Merget *et al.*, 2017). These five comparison models use the molecular structure feature of each compound as their input and whether the kinase is associated with the compound as a constraint to train the model for predicting potential kinase inhibitors. That is, these prediction models are local prediction models, which may not be sufficient for checking the performance of the model at the global level. Therefore, to verify the global performance of the IDDkin model, we added several comparison methods commonly used for association prediction on complex networks in the field of bioinformatics and virtual drug screening: Random Walk with Restart (RWR) (Lv *et al.*, 2015; Tong *et al.*, 2006), Non-negative Matrix Factorization (NMF) (Lee and Seung, 1999; 2001) and Katz (Chen *et al.*, 2017; Katz, 1953). All of the above local comparison methods and global comparison methods used the data given in Table 1. A detailed introduction and the parameter selection of these comparison methods are shown in the Supplementary Materials.

3.3 Implementation and evaluation strategy

We evaluated the performance of IDDkin and baselines by performing 5-fold cross-validation for the prediction of kinase inhibitors. All known compound-kinase pairs were randomly divided into five equal sets, four of which were used to train the prediction model, while the remaining set was used as the test set. In this study, we

used the same implementation of IDDkin for two kinase profiling datasets. We implemented IDDkin with TensorFlow 1.0 (Abadi *et al.*, 2016) and used Adam for training with a learning rate of 0.0005. We chose the dimension of kinase free embedding L from {8, 16, 32, 64, 12}, the dimension p of the projection matrix G_p and G_Q from {8, 12, 16, 20, 24} and the parameter K of the K-nearest-neighbour from {6, 8, 10, 12, 14}.

The area under the precision-recall curve (AUPR) and the area under the receiver operating characteristic curve (AUC) were used to evaluate the overall performance of IDDkin. To evaluate the performance of the model more comprehensively, we also used balanced accuracy, precision, recall and the F1 score to verify the performance of the model. The detailed calculation process of these evaluation metrics is shown in the Supplementary Materials. To reduce the bias of cross-validation, we performed the whole process ten times and took the average value as the final result.

3.4 Comparison with previously reported methods

We carried out a parallel comparison with previously reported methods to verify the predictive performance and generalizability of our IDDkin model. In these comparison methods [including MTDNN (Li *et al.*, 2019), PEPECFP (Avram *et al.*, 2018), KNN (Schurer and Muskal, 2013), Naïve Bayes (Schurer and Muskal, 2013) and Merget's method (Merget *et al.*, 2017)], the molecular structure characteristics of the compound are used as the input, and whether the compound is related to the kinase is used as a constraint to train the model. Afterward, we also used 5-fold cross-validation to test the performance of these models. Unlike the 5-fold cross-validation at the global level, we randomly divided the compounds associated with each kinase into five equal parts, four of which are used for training and the remaining one is used for testing. It is worth noting that both the IDDkin model and the comparison methods should be retrained at each fold in the 5-fold cross-validation. The obvious problem here is that if some kinases are only associated with a small number of compounds, then these models are prone to overfitting when predicting potential kinase inhibitors. Inspired by Merget's method (Merget *et al.*, 2017), we only consider datasets with at least 20 bioactivity values per kinase in the Tang set. Since the data of the PKIS set are small and sparse, we only consider datasets with at least 10 bioactivity values per kinase in the PKIS set. We collected the AUC and AUPR of each kinase in each method to draw a box plot, as shown in Figure 2 below. In Figure 2A, the average AUCs of IDDkin in both datasets were greater than 0.9 and higher than those of the comparison methods. Figure 2B shows the AUPR of IDDkin and the comparison methods. It is worth noting that the average AUPR of IDDkin in the PKIS set was close to 0.4, which is much higher than the comparison methods, indicating the superior performance of the IDDkin model in the prediction of kinase inhibitors. The above results demonstrate that IDDkin is a feasible model for predicting kinase inhibitors. Furthermore, to identify whether the presented differences between the compared models are significant, we used a paired *t*-test via 5-fold cross-validation to compare the AUCs and AUPRs of IDDkin and the other methods. As shown in Table 2, the *P*-values were less than 0.05, indicating that the differences between the AUCs and AUPRCs are statistically significant.

Supplementary Figure S1 demonstrates that even though the number of training bioactivity data points for the kinases is quite small, IDDkin shows decent performance. On the Tang dataset, given a kinase with fewer than 25 training data points, the performance of IDDkin is much better than the comparison methods. The same conclusion could be reached for the PKIS dataset. In particular, the AUPR metric of IDDkin is much higher than those of the comparison methods. Similarly, IDDkin also shows better performance than the comparison methods when the number of training data points of a kinase is large. These results show that the IDDkin model has good generalizability; that is, regardless of the number of training data points, this model can always achieve better prediction performance.

Table 1. Sizes of final datasets used for the prediction of kinase inhibitors

	Compounds	Kinases	Positive	Negative	Density
Tang	1351	188	15660	238328	6.17%
PKIS	366	195	2414	68956	3.38%

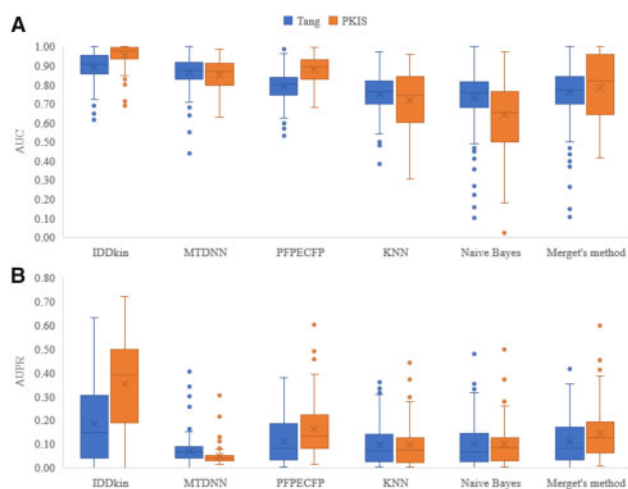


Fig. 2. Performance comparison between IDDkin and other comparison methods. (A) Comparison of the performance of IDDkin model and comparison methods with AUC as evaluation metric in Tang set and PKIS set. (B) Comparison of the performance of IDDkin model and comparison methods with AUPR as evaluation metric in Tang set and PKIS set

Table 2. *P*-values obtained through paired *t*-test of the AUCs and AUPRCs of IDDkin and other comparative methods

		MTDNN	PFPECFP	KNN	Naïve Bayes	Merget's method
Tang	AUCs	4.30E-4	1.08E-33	1.82E-41	3.50E-35	1.36E-26
	AUPRs	5.75E-20	6.27E-08	3.24E-10	2.80E-10	1.81E-08
PKIS	AUCs	1.44E-15	4.12E-11	3.36E-26	8.49E-31	4.67E-14
	AUPRs	3.38E-41	3.13E-11	2.85E-18	4.96E-20	3.2E-14

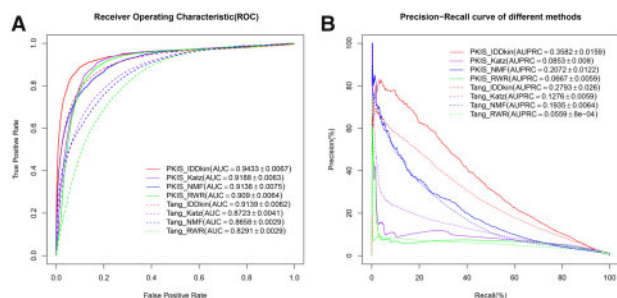


Fig. 3. Performance of different methods on the constructed compound-kinase network. (A) Receiver operating characteristic (ROC) curves of IDDkin and comparison methods using PKIS set and Tang set with 5-fold cross-validation. (B) Precision-recall (PR) curves of IDDkin and comparison methods using PKIS set and Tang set with 5-fold cross-validation

3.5 The global predictive performance evaluation of IDDkin

To verify the global performance of IDDkin in the prediction of kinase inhibitors, The area under the receiver operating characteristic curve (AUC) and the area under the precision-recall curve (AUPR) were utilized to verify the overall performance of IDDkin. We found that IDDkin achieves a higher accuracy in the 5-fold cross-validation than the state-of-the-art methods (Table 4). In the PKIS dataset, the AUC value of IDDkin was 0.9433, which is 2.67% higher than the second-ranked Katz model (AUC=0.9188) (Fig. 3A). For another evaluation metric AUPR, the IDDkin model achieved 0.3582 in the PKIS dataset, which is 72.88% higher than the second-ranked NMF model (AUPR = 0.2072) (Fig. 3B). Whether it is on the PKIS dataset or the Tang dataset, the AUPR values of the IDDkin model and the comparison method are low, which

Table 3. *P*-values obtained through paired *t*-test of the AUCs and AUPRCs of IDDkin and other compared methods for 10 runs

	Tang			PKIS		
	Katz	NMF	RWR	Katz	NMF	RWR
AUCs	3.25E-09	1.7E-08	1E-11	9.86E-05	4.95E-06	4.13E-07
AUPRs	3.48E-08	1.52E-04	2.11E-09	3.55E-12	5.19E-09	2.72E-12

is caused by the imbalance of the positive and negative samples of the two datasets. In general, IDDkin performs best among all methods both on the PKIS dataset and the Tang dataset. These results indicate that the IDDkin model has excellent global prediction capability. Furthermore, the AUCs and AUPRs of IDDkin and the comparison methods with different runs were compared using paired *t*-tests (two-tailed tests) via 5-fold cross-validation. As shown in Table 3, the *P*-values were less than 0.05, suggesting that the differences between AUCs and AUPRs are statistically significant.

The number of correctly predicted true positives reflects the discriminability of a prediction model to distinguish true positives, especially when the number of negative samples is much larger than the number of positive samples (Zeng et al., 2019). Therefore, we added four additional evaluation metrics, recall, precision, F1 score and balanced accuracy, to verify the performance of the various methods. As shown in Table 4, both on the Tang dataset and the PKIS dataset, the prediction accuracy of IDDkin is higher than those of the comparison methods. It is worth noting that the prediction performance improvement of IDDkin on the Tang set is more significant than that of the PKIS set. For example, compared with the Katz model on the Tang set, the performance of IDDkin in terms of recall, precision, F1 score and balanced accuracy were improved by 0.0411, 0.0093, 0.017 and 0.0011, respectively. However, compared with the Katz model on the PKIS set, the performance of

Table 4. The performance comparison of IDDkin and other models using multiple evaluation metrics on two datasets

	Tang				PKIS			
	Recall	Precision	F1-score	Balanced accuracy	Recall	Precision	F1-score	Balanced accuracy
IDDkin	0.9085	0.0482	0.0915	0.7052	0.9403	0.0314	0.0608	0.7214
Katz	0.8674	0.0389	0.0745	0.6831	0.9159	0.0233	0.0454	0.7098
NMF	0.8610	0.0414	0.0790	0.6795	0.9108	0.0269	0.0523	0.7066
RWR	0.8248	0.0308	0.0594	0.6645	0.9061	0.0216	0.0422	0.7047

Table 5. Metrics of IDDkin with different numbers of layers

	Layers	AUC	AUPR	Balanced accuracy	Recall	Precision	F1-score
	2	0.9139	0.2793	0.7052	0.9085	0.0482	0.0915
	3	0.9092	0.2662	0.7046	0.9039	0.0474	0.0901
	4	0.9073	0.2875	0.7025	0.9021	0.0481	0.0913
PKIS	1	0.9297	0.3129	0.7126	0.9267	0.0301	0.0583
	2	0.9433	0.3582	0.7214	0.9403	0.0314	0.0607
	3	0.9371	0.3542	0.7147	0.9341	0.0312	0.0604
	4	0.9315	0.3476	0.7192	0.9285	0.0309	0.0598

IDDkin in terms of recall, precision, F1 score and balanced accuracy were only improved by 0.0244, 0.0081, 0.0154 and 0.0003, respectively. Judging from the comparison between the IDDkin and Katz models on four verification metrics, the performance of IDDkin on the PKIS set has been more significantly improved than the performance of IDDkin. This conclusion not only appears in the comparison with Katz but also in comparison with the other two models. In fact, the size of the Tang dataset is larger than that of the PKIS dataset, and its number of data points is 6.5 times that of the PKIS set. Therefore, we infer that the IDDkin model may have more significant performance improvements on large datasets. As more kinases are discovered, IDDkin may have a greater performance improvement in predicting kinase inhibitors in the future. There are maybe two main reasons for the high accuracy of the IDDkin model: (i) The heterogeneous network comprised of compounds and kinases integrates multiple sources of information, including compound similarity information and kinase-compound association information; and (ii) The three methods of graph convolutional network, graph attention mechanism and adaptive weighting fully mine the heterogeneous network information and improve the performance of the model.

3.6 Detailed model analysis

We analysed the depth l of the influence deep diffusion model, also known as the number of layers of GCN used to aggregate information. Table 5 shows the results of IDDkin with different layers. When the number of layers is 1, the GCN of the IDDkin model only fuses the information of the first-order neighbours, and its performance is not very high. As can be observed from this table, when we leverage the layerwise fusion process from layer = 1 to layer = 2, the performance increases quickly for both datasets. For the Tang and PKIS sets, when the number of GCN fusion layers is 2, the IDDkin model achieves the best performance. As the number of layers increases to 3, the performance of IDDkin begins to drop. This trend of performance degradation becomes more obvious as the number of layers increases. Since the influence between the compounds decreases while the distance between each compound and the neighbour increases, setting the number of layers to 2 is sufficient to predict the kinase inhibitors. In fact, there is a similar conclusion when applying deep GCN in the recommendation system, and the number of layers is generally set to 2 or 3 (Kipf and Welling, 2017; Ying *et al.*, 2018).

We also present a sensitivity analysis for other parameters in IDDkin, including the dimension L of kinase free embedding, the dimension p of the projection matrix and the parameter K of K -nearest neighbour. In Supplementary Figure S2(A), the AUC and AUPR of IDDkin gradually increase when embedding dimension L increases from 8 to 32. Generally, a reference value $L = 32$ gives optimal results. As L continues to increase, the accuracy of the IDDkin model decreases, which may be due to excessive dimensionality that easily introduces noise. The features of the compound and the embedding of the kinase are projected into the space of dimension p through the projection matrix for reconstructing the original association matrix. As shown in Supplementary Figure S2(B), the dimension p could affect the performance of IDDkin. It is inappropriate for the dimension of the projection matrix to be too large or too small. Too small may lose effective information, and too large may easily contain noisy data. Typically, a balanced choice such as $p = 12$ is desirable. Supplementary Figure S2(C) shows the impact of the K -nearest-neighbour parameter K . It can be seen from Supplementary Figure S2C that IDDkin can achieve the best results when $K = 10$. The values of each inflection point in Supplementary Figure S2(A)–(C) are listed in Supplementary Tables S1–S3.

We present some quantitative convergence analysis results to verify that the objective function of our IDDkin is non-increasing. In this study, the universal applicability of our method was evaluated through two datasets (i.e. the PKIS set and the Tang set) and three metrics (i.e. the loss, AUC and AUPR) as examples. As shown in Supplementary Figure S3(A), the training loss quickly drops within the first 1000 epochs and gradually begins to converge in ~ 1500 epochs. Supplementary Figure S3(B) and (C) show the changes in AUC and AUPR with the number of IDDkin model iterations. We can draw the same conclusion from Supplementary Figure S3, that is, the three metrics on the test data also begin to converge in ~ 1500 epochs. It is worth noting that as the number of iterations increases, the performance of the model tends to decline, which may be due to the slight overfitting of the model. In summary, the loss between consecutive feature representation matrices is non-increasing, and our IDDkin model converges very fast.

3.7 Performance of IDDkin by ablation study

To investigate how the deep GCN, GAT and adaptive weighting improve the performance of the proposed model, the following variants of IDDkin were conducted for the ablation study:

IDDkin-GCN is a variant model that removes the Deep GCN in IDDkin and directly uses the original feature vector of the compounds as their representation. The information diffused to the kinase at the GAT stage is only the original structural feature information of each compound, which does not contain its neighbouring information.

IDDkin-GAT replaces the GAT in IDDkin with adaptive weighting to verify the importance of the GAT part to the model.

IDDkin-AW directly removes the adaptive weighting part, that is, the final compound representation vector does not contain compound-kinase pair association information.

Table 6 shows the performance of IDDkin with different variant models. As seen from this figure, the overall performance will be reduced to varying degrees when removing or replacing some important components in the IDDkin model. From the perspective of the three variant models, the performance of the IDDkin-GAT model has the smallest overall decrease. However, this does not

Table 6. AUC and AUPR of IDDkin's variant models on PKIS set and Tang set for ablation study

Simplified models	Tang				PKIS			
	AUC	Improve	AUPR	Improve	AUC	Improve	AUPR	Improve
IDDkin	0.9139	–	0.2793	–	0.9433	–	0.3582	–
IDDkin-GCN	0.8919	–2.41%	0.1934	–30.76%	0.9379	–0.57%	0.2774	–22.56%
IDDkin-GAT	0.9109	–0.33%	0.2654	–4.98%	0.9392	–0.43%	0.3237	–9.63%
IDDkin-AW	0.8343	–8.71%	0.2378	–14.86%	0.9224	–2.22%	0.3347	–6.56%

mean that the compound-kinase association information is not important for the performance improvement of the model because we still consider the association information in IDDkin-GAT and only use adaptive weighting instead. The performance changes of IDDkin-AW can also reflect the importance of related information for the IDDkin model. Especially in the Tang dataset, the AUC value of the IDDkin-AW model decreases by 8.71%, and the AUPR value decreases by 14.84%. Since the IDDkin-AW model removes the adaptive weighting part, the representation vector of the compound does not contain related information. Therefore, the performance of the IDDkin-AW model is seriously degraded, which fully illustrates the importance of the related information in the compound-kinase pair for improvement of the model performance. The deep GCN also has an important impact on the performance of IDDkin. The AUPR value of the IDDkin-GCN model on the Tang set and the PKIS set decreases by 30.76% and 22.56% compared to the IDDkin model. This shows that deep GCN is also the key to improving the performance of the IDDkin model. Therefore, each part of IDDkin, including GCN, GAT and adaptive weighting, is very important for the prediction of kinase inhibitors.

3.8 Generalization ability of IDDkin

To further demonstrate the actual potential for predicting kinase inhibitors of IDDkin, we performed an additional experiment based on the new test data. In the Tang and PKIS sets, kinases are divided into the inactive (negative) or active (positive) class with the pIC_{50} cut-off set to 6.0. We used the data points with a pIC_{50} cut-off between 6.0 and 6.3 as the test set and the remaining active data points as the training set. To reduce the data bias of the overall process, we repeated it 10 times and obtained the average performance. As shown in [Supplementary Figure S4](#), IDDkin shows better accuracy than state-of-the-art methods on the PKIS set and Tang set. Furthermore, the AUCs and AUPRs of IDDkin and the comparison methods with 10 runs were compared using paired *t*-tests (two-tailed tests). As shown in [Supplementary Figure S4](#), IDDkin (AUC = 0.8860; AUPR = 0.2615) shows better accuracy than the state-of-the-art methods (Katz: AUC = 0.8793, AUPR = 0.1341; NMF: AUC = 0.8720, AUPR = 0.2351; RWR: AUC = 0.8545, AUPR = 0.0929) on the PKIS set, and there are similar conclusions for the Tang set. Furthermore, the AUCs and AUPRs of IDDkin and the comparison methods with 10 runs were compared using paired *t*-tests. As shown in [Supplementary Table S4](#), the *P*-values were less than 0.05, suggesting that the differences between AUCs and AUPRs are statistically significant. In addition, four additional evaluation metrics (recall, precision, F1 score and balanced accuracy) were employed to verify the superiority of IDDkin. As shown in [Supplementary Table S5](#), for both the Tang dataset and the PKIS dataset, the prediction accuracy of IDDkin is higher than that of the comparison methods. For example, the recall of the IDDkin model on the Tang set is 0.7845, which is 1.73%, 3.37% and 8.22% higher than those of the Katz, NMF and RWR methods, respectively. The IDDkin model also has obvious advantages in the PKIS dataset and the other evaluation metrics. These results indicate that the IDDkin model can achieve good performance under different test data, indicating its strong generalization ability.

3.9 Case studies: sorafenib, dasatinib and sunitinib

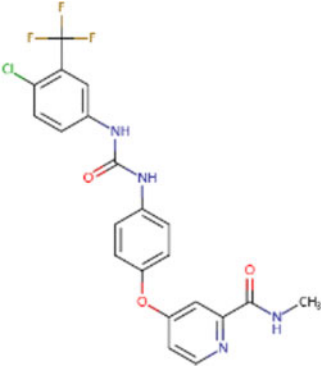
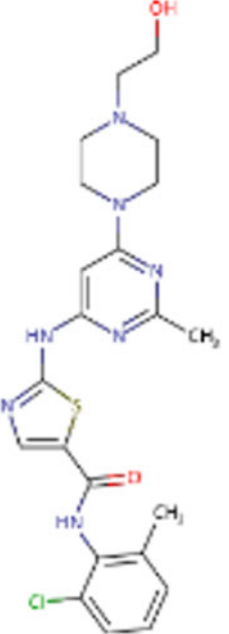
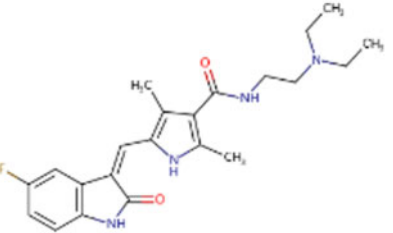
Three anticancer drugs approved by the US FDA, sorafenib ([Wilhelm et al., 2006](#)), dasatinib ([Talpez et al., 2006](#)) and sunitinib ([Motzer et al., 2007](#)), were selected from the compound data as the objects of the case studies. These drugs are all FDA approved and are used to treat cancers, including renal cell carcinoma, hepatocellular carcinoma, haematological malignancies, pancreatic neuroendocrine tumours, renal cancer and gastrointestinal stromal tumours. Sorafenib ([Wilhelm et al., 2006](#)) is a US-FDA-approved anticancer drug for the treatment of advanced renal cell carcinoma and non-resectable hepatocellular carcinoma. Dasatinib ([Talpez et al., 2006](#)) is an US-FDA-approved orally administered drug that is used to treat certain haematological malignancies. Sunitinib ([Motzer et al., 2007](#)) is approved by the US FDA for the treatment of pancreatic neuroendocrine tumours, renal carcinoma and imatinib-resistant gastrointestinal stromal tumours (GISTs). A description of these three anticancer drugs is given in the second column of [Table 7](#). For each drug, we selected 8 kinases with the highest predictive correlation scores and checked whether the predicted results were verified by other literature in PubMed.

The experimental results are shown in [Table 7](#). Six, five and four candidate kinases for sorafenib, dasatinib and sunitinib, respectively, were identified and supported by direct evidence. For example, the literature ([Luo et al., 2020a,b](#)) clearly indicates that s-HBEGF regulates sorafenib-induced HFSR by relying on JNK2 to stabilize SIRT1. Furthermore, a study of Src kinase inhibitors ([Ren et al., 2011](#)) mentioned that the Src kinase inhibitor dasatinib kills established murine leukaemia cell lines expressing chimeric FGFR1 kinases. In summary, the results of these case studies have demonstrated that IDDkin is a useful tool for predicting potential kinase inhibitors.

4 Discussion and conclusion

Previous studies have found that graph neural network methods achieve significantly better prediction accuracy than traditional models ([Quan et al., 2019](#); [Zeng et al., 2019, 2020](#)) in drug discovery studies due to adequate and valid network knowledge in heterogeneous networks. In this study, we developed a novel network-based influence deep diffusion model (IDDkin) to uncover the bioactivities of small molecules at a kinome-wide level. Apart from the gold standard compound-kinase association network, we incorporated a compound similarity network to construct a complicated heterogeneous network that provides a multiperspective view and diverse information for enhancing the prediction of kinase inhibitors. IDDkin first uses a deep graph convolution network to aggregate the high-order neighbour information in the compound similarity network and then diffuses the information in the updated compound node to the kinase node through the graph attention network. Finally, the updated kinase nodes will diffuse the influence to the compound nodes by the adaptive weighting method again so that the compound nodes contain both similarity network knowledge and related network information. Theoretically, IDDkin is superior to existing prediction methods of kinase inhibitors because we incorporated topological information into the compound similarity network and the compound-kinase association network. Furthermore, IDDkin uses the K-nearest neighbour method to eliminate low-similarity associations in the similarity network and considers the higher-order topology information in the compound

Table 7. The top eight potential kinase candidates detected by IDDkin based on PubMed for the three selected compounds

Compounds	Description	Kinases	Groups	Score	Evidence
 <p>Sorafenib</p>	<p>Sorafenib (Wilhelm <i>et al.</i>, 2006) is a US-FDA-approved anticancer drug for the treatment of advanced renal cell carcinoma and non-resectable hepatocellular carcinoma.</p>	JNK2	CMGC	0.5449	32296111
		BLK	TK	0.5280	-
		FRK	TK	0.4959	22500798
		LYN	TK	0.4904	26912052
		AurA	Other	0.3830	-
		EphA2	TK	0.3809	32203105
		FYN	TK	0.3647	28625738
 <p>Dasatinib</p>	<p>Dasatinib (Talpoz <i>et al.</i>, 2006) is an US-FDA-approved orally administered, which used to the treatment of certain hematological malignancies</p>	CK1d	CK1	0.3631	-
		FGFR1	TK	0.5962	21937681
		SLK	STE	0.4755	-
		JNK2	CMGC	0.4724	29266867
		FGFR3	TK	0.4348	32370101
		HGK	STE	0.4266	-
		NEK2	Other	0.4187	23652925
		MET	TK	0.3996	30482914
 <p>Sunitinib</p>	<p>Sunitinib (Motzer <i>et al.</i>, 2007) is approved by US-FDA for the treatment of pancreatic neuroendocrine tumours, renal carcinoma and imatinib-resistant gastrointestinal stromal tumour (GIST).</p>	KDR	TK	0.3905	-
		GCK	STE	0.9097	-
		FRK	TK	0.8161	-
		AurA	Other	0.8135	-
		TRKC	TK	0.7489	26009590
		FER	TK	0.7120	-
		TRKB	TK	0.7084	24759734
MNK2	CAMK	0.7033	19844230		
Sunitinib		FYN	TK	0.7002	27589830

similarity network. The use of the graph attention network and adaptive weighting method fully mines the effective information about compound-kinase pairs, which helps to improve the performance of the model. To verify these points, we used 5-fold cross-validation to confirm IDDkin's predictive ability. The IDDkin model achieves superior prediction performance compared with the traditional kinase inhibitor prediction model or the models commonly used for global association prediction based on a heterogeneous network. IDDkin makes full use of the structural information of the compounds and the heterogeneous network information, while the traditional kinase inhibitor prediction models only use the structural information of the compound and association information [i.e. MTDNN (Li et al., 2019), PFPECFP (Avram et al., 2018)]. The models commonly used in association prediction only consider heterogeneous network information and do not incorporate compound structure information [i.e. NMF (Lee and Seung, 1999; Lee and Seung, 2001), RWR (Lv et al., 2015; Tong et al., 2006)], which may be an important reason why their global prediction is worse than that of IDDkin. In addition, we also studied the importance of each stage of the IDDkin model to improve the performance of the model through an ablation study. To verify the generalizability of IDDkin, we lowered the activation threshold to $pIC_{50} = 6.0$ in the Tang and PKIS sets and used the data points with a pIC_{50} cut-off between 6.0 and 6.3 as the test set. Finally, we conducted a case study to confirm the IDDkin model's ability to predict new compound-kinase associations. In these experiments, IDDkin achieved good performance, confirming its strong predictive ability and generalization ability.

In future work, since IDDkin is an extensible framework, merging or collecting more effective information from the database or literature may improve the model's predictive ability. In addition, due to the limitations of the deep influence diffusion model, IDDkin can only incorporate compound-related information at present. The direction for future development is to modify the structure of the deep influence diffusion model and extend IDDkin to integrate both compound-related information and kinase-related information, which include the sequence information and the functional information of the kinases. In summary, the superior predictive performance of our model is attributed to compound-kinase heterogeneous networks, network-based deep learning and graph neural network methods. From a translational perspective, the network tools developed here can help develop novel and effective drug discovery methods to treat many complex diseases from the perspective of network-based kinase inhibitor prediction if broadly applied.

Funding

This work was supported by the National Natural Science Foundation of China [61873089, 62002154], Hunan Provincial Innovation Foundation for Postgraduate [CX20200436] and Scientific Research Startup Foundation of University of South China [190XQD096].

Conflict of Interest: none declared.

References

- Abadi, M. et al. (2016) Tensorflow: a system for large-scale machine learning. *OSDI*, **16**, 265–283.
- Anastasiadis, T. et al. (2011) Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1039–1045.
- Avram, S. et al. (2018) Modeling kinase inhibition using highly confident data sets. *J. Chem. Inf. Model.*, **58**, 957–967.
- Barabasi, A. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.
- Bhullar, K.S. et al. (2018) Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol. Cancer*, **17**, 1–20.
- Bleakley, K. and Yamanishi, Y. (2009) Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics*, **25**, 2397–2403.
- Bora, A. et al. (2016) Predictive models for fast and effective profiling of kinase inhibitors. *J. Chem. Inf. Model.*, **56**, 895–905.

- Cao, D. et al. (2013) Large-scale prediction of human kinase–inhibitor interactions using protein sequences and molecular topological structures. *Anal. Chim. Acta*, **792**, 10–18.
- Chen, X. et al. (2017) A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics*, **33**, 733–739.
- Cheng, F. et al. (2018) Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.*, **9**, 2691–2691.
- Cheng, F. et al. (2019) Network-based prediction of drug combinations. *Nat. Commun.*, **10**, 1197–1208.
- Davis, M.I. et al. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1046–1051.
- Dickson, M. and Gagnon, J.P. (2004) The cost of new drug discovery and development. *Discov. Med.*, **4**, 172–179.
- Ding, P. et al. (2020) Incorporating multisource knowledge to predict drug synergy based on graph co-regularization. *J. Chem. Inf. Model.*, **60**, 37–46.
- Ding, P. et al. (2019) Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological and network knowledge. *IEEE J. Biomed. Health Inf.*, **23**, 1336–1345.
- Du, Y. et al. (2020) Energy-based models for atomic-resolution protein conformations. In: *International Conference on Learning Representations*. Addis Ababa, Ethiopia.
- Elkins, J.M. et al. (2016) Comprehensive characterization of the published kinase inhibitor set. *Nat. Biotechnol.*, **34**, 95–103.
- Fabbro, D. et al. (2015) Ten things you should know about protein kinases: IUPHAR Review 14. *Br. J. Pharmacol.*, **172**, 2675–2700.
- Gilson, M.K. et al. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–D1053.
- Janssen, A.P.A. et al. (2019) Drug discovery maps, a machine learning model that visualizes and predicts kinome-inhibitor interaction landscapes. *J. Chem. Inf. Model.*, **59**, 1221–1229.
- Katz, L. (1953) A new status index derived from sociometric analysis. *Psychometrika*, **18**, 39–43.
- Kipf, T.N. and Welling, M. (2017) Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations*. Toulon, France.
- Knapp, S. et al. (2013) A public-private partnership to unlock the untargeted kinome. *Nat. Chem. Biol.*, **9**, 3–6.
- Kong, Y. and Yu, T. (2020) forgeNet: a graph deep neural network model using tree-based ensemble classifiers for feature graph construction. *Bioinformatics*, **36**, 3507–3515.
- Lee, D.D. and Seung, H.S. (1999) Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791.
- Lee, D.D. and Seung, H.S. (2001) Algorithms for non-negative matrix factorization. In: *Advances in Neural Information Processing Systems*. Vancouver, Canada, pp. 556–562.
- Li, X. et al. (2019) Deep learning enhancing kinome-wide polypharmacology profiling: model construction and experiment validation. *J. Med. Chem.*, **63**, 8723–8737.
- Long, Y. et al. (2020) Predicting human microbe–drug associations via graph convolutional network with conditional random field. *Bioinformatics*, **36**, 4918–4927.
- Luo, J. et al. (2020a) Incorporating clinical, chemical and biological information for predicting small molecule–microRNA associations based on non-negative matrix factorization. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, In press. DOI: 10.1109/TCBB.2020.2975780
- Luo, J. et al. (2020b) s-HBEGF/SIRT1 circuit-dictated crosstalk between vascular endothelial cells and keratinocytes mediates sorafenib-induced hand–foot skin reaction that can be reversed by nicotinamide. *Cell Res.*, **30**, 779–715.
- Luo, Y. et al. (2017) A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.*, **8**, 573–573.
- Lv, Y. et al. (2015) Identifying novel associations between small molecules and miRNAs based on integrated molecular networks. *Bioinformatics*, **31**, 3638–3644.
- Manallack, D.T. et al. (2002) Selecting screening candidates for kinase and G protein-coupled receptor targets using neural networks. *J. Chem. Inf. Comput. Sci.*, **42**, 1256–1262.
- Manning, G. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
- Merget, B. et al. (2017) Profiling prediction of kinase inhibitors: toward the virtual assay. *J. Med. Chem.*, **60**, 474–485.
- Metz, J.T. et al. (2011) Navigating the kinome. *Nat. Chem. Biol.*, **7**, 200–202.

- Moret, M. *et al.* (2020) Generative molecular design in low data regimes. *Nat. Mach. Intell.*, **2**, 171–180.
- Motzer, R.J. *et al.* (2007) Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N. Engl. J. Med.*, **356**, 115–124.
- Natarajan, N. and Dhillon, J.S. (2014) Inductive matrix completion for predicting gene–disease associations. *Bioinformatics*, **30**, i60–i68.
- Nguyen, P.A. *et al.* (2019) Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat. Commun.*, **10**, 1–11.
- Niijima, S. *et al.* (2012) Dissecting kinase profiling data to predict activity and understand cross-reactivity of kinase inhibitors. *J. Chem. Inf. Model.*, **52**, 901–912.
- Noble, M.E.M. *et al.* (2004) Protein kinase inhibitors: insights into drug design from structure. *Science*, **303**, 1800–1805.
- Pittala, S. and Bailey-Kellogg, C. (2020) Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinformatics*, **36**, 3996–4003.
- Quan, Z. *et al.* (2019) Graphcpi: Graph neural representation learning for compound–protein interaction. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, San Diego, California, USA, pp. 717–722.
- Ren, M. *et al.* (2011) Src activation plays an important key role in lymphoma-genesis induced by FGFR1 fusion kinases. *Cancer Res.*, **71**, 7312–7322.
- Roskoski, R. Jr, (2020) Properties of FDA-approved small molecule protein kinase inhibitors: a 2020 update. *Pharmacol. Res.*, **152**, 104609.
- Schurer, S.C. and Muskal, S.M. (2013) Kinome-wide activity modeling from diverse public high-quality data sets. *J. Chem. Inf. Model.*, **53**, 27–38.
- Shen, C. *et al.* (2020a) Multiview joint learning-based method for identifying small-molecule-associated miRNAs by integrating pharmacological, genomics, and network knowledge. *J. Chem. Inf. Model.*, **60**, 4085–4097.
- Shen, C. *et al.* (2020b) Identification of small molecule–miRNA associations with graph regularization techniques in heterogeneous networks. *J. Chem. Inf. Model.* **60**, 6709–6721.
- Talpaz, M. *et al.* (2006) Dasatinib in imatinib-resistant Philadelphia chromosome–positive leukemias. *N. Engl. J. Med.*, **354**, 2531–2541.
- Tang, J. *et al.* (2014) Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J. Chem. Inf. Model.*, **54**, 735–743.
- Tong, H. *et al.* (2006) Fast random walk with restart and its applications. In: *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, Hong Kong, China, pp. 613–622.
- Tsubaki, M. *et al.* (2019) Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, **35**, 309–318.
- Veličković, P. *et al.* (2018) Graph attention networks. In: *International Conference on Learning Representations*. Vancouver, Canada.
- Wan, F. *et al.* (2019) NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*, **35**, 104–111.
- Wang, H. *et al.* (2020) GoGNN: graph of graphs neural network for predicting structured entity interactions. In: *International Joint Conference on Artificial Intelligence*. Springer, Yokohama, Japan.
- Wilhelm, S. *et al.* (2006) Discovery and development of sorafenib: a multi-kinase inhibitor for treating cancer. *Nat. Rev. Drug Discov.*, **5**, 835–844.
- Wu, L. *et al.* (2019) A neural influence diffusion model for social recommendation. In: *International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris, France, pp. 235–244.
- Xuan, P. *et al.* (2019) Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics*, **35**, 4108–4119.
- Yabuuchi, H. *et al.* (2011) Analysis of multiple compound–protein interactions reveals novel bioactive molecules. *Mol. Syst. Biol.*, **7**, 472–472.
- Ying, R. *et al.* (2018) Graph convolutional neural networks for web-scale recommender systems. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, London, United Kingdom, pp. 974–983.
- Zang, C. and Wang, F. (2020) MoFlow: an invertible flow model for generating molecular graphs. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Conference virtual platform (VFairs), pp. 617–626.
- Zeng, X. *et al.* (2020) Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics*, **36**, 2805–2812.
- Zeng, X. *et al.* (2019) deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics*, **35**, 5191–5198.
- Zhang, Z. *et al.* (2020) A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. *Bioinformatics*, **36**, 3474–3481.