Journal of Cheminformatics

RESEARCH

Open Access

Large-scale comparison of machine learning methods for profiling prediction of kinase inhibitors

Jiangxia Wu[1†], Yihao Chen[1†], Jingxing Wu[1†], Duancheng Zhao[1], Jindi Huang[1], MuJie Lin[1] and Ling Wang[1*]

Abstract

Conventional machine learning (ML) and deep learning (DL) play a key role in the selectivity prediction of kinase inhibitors. A number of models based on available datasets can be used to predict the kinase profile of compounds, but there is still controversy about the advantages and disadvantages of ML and DL for such tasks. In this study, we constructed a comprehensive benchmark dataset of kinase inhibitors, involving in 141,086 unique compounds and 216,823 well-defined bioassay data points for 354 kinases. We then systematically compared the performance of 12 ML and DL methods on the kinase profiling prediction task. Extensive experimental results reveal that (1) Descriptor-based ML models generally slightly outperform fingerprint-based ML models in terms of predictive performance. RF as an ensemble learning approach displays the overall best predictive performance. (2) Single-task graph-based DL models are generally inferior to conventional descriptor- and fingerprint-based ML models, however, the corresponding multi-task models generally improves the average accuracy of kinase profile prediction. For example, the multi-task FP-GNN model outperforms the conventional descriptor- and fingerprint-based ML models with an average AUC of 0.807. (3) Fusion models based on voting and stacking methods can further improve the performance of the kinase profiling prediction task, specifically, RF::AtomPairs + FP2 + RDKitDes fusion model performs best with the highest average AUC value of 0.825 on the test sets. These findings provide useful information for guiding choices of the ML and DL methods for the kinase profiling prediction tasks. Finally, an online platform called KIPP (https://kipp.idruglab.cn) and python software are developed based on the best models to support the kinase profiling prediction, as well as various kinase inhibitor identification tasks including virtual screening, compound repositioning and target fishing.

Keywords Kinase profiling, Machine learning, Deep learning, Molecular fingerprints, Molecular graphs

Introduction

The human kinome comprises more than 500 kinases, constituting approximately 1.7% of all human genes [1]. Protein kinases (PKs) play central roles in mediating most signaling pathways involved in cellular metabolism, transcription, cell cycle, apoptosis, and differentiation. Therefore, PKs have become one of the most interesting classes of drug targets for various diseases, including cancers [2–4], inflammation [5, 6], central nervous system disorders [7], cardiovascular diseases [8], complications of diabetes [9], and Alzheimer's disease [10]. As

†Jiangxia Wu, Yihao Chen and Jingxing Wu have contributed equally.

*Correspondence:
Ling Wang
lingwang@scut.edu.cn
[1] Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, Joint International Research Laboratory of Synthetic Biology and Medicine, Guangdong Provincial Engineering and Technology Research Center of Biopharmaceuticals, School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

such a significant class of targets, kinase inhibitors have been the focus of drug discovery. There are currently 71 FDA-approved small-molecule kinase inhibitors. In addition, approximately 110 innovative kinases are emerging as targets for drugs development in clinical trials [11]. Most FDA-approved drugs (63/71) targeting kinases are ATP-competitive inhibitors which inhibit kinases activity by binding to the ATP binding site of the kinase domain. However, the intrinsically highly conserved ATP binding sites of kinases may lead to off-target effects (i.e., low selectivity) of kinase inhibitors, potentially leading to undesirable side effects. Accordingly, identifying selective PK inhibitors remains an important challenge in the development of kinase-targeted drugs. Traditional kinase inhibitor assays are low-throughput methods that primarily measure the ability of compounds to reduce the phosphorylation activity for a given kinase (e.g. IC_{50}) or their binding affinities to a kinase (dissociation constant, such as K_i and K_d). Notably, such measurement methods typically do not extend to the ability of a compound to inhibit the entire kinome. High-throughput kinase profiling assay has also become feasible in recent years, but the excessive cost makes it difficult to use as a routine early stage of drug discovery efforts [12].

Based on experimental data, a number of computational methods have been developed and published elsewhere, aiming to significantly reduce the cost, time and laborious involved in experimental identification. Generally, these computational methods can be classified into two major categories: structure- and ligand-based kinase inhibition and/or profiling prediction approaches (called virtual assay). Molecular docking, commonly used in structure-based prediction methods for kinase inhibition, has good generalizability, but its accuracy depends on the crystal structure of the kinase and the accuracy of the scoring function [13, 14]. Ligand-based methods include pharmacophore modelling, and quantitative structure–activity relationship (QSAR) [15–21]. Based on different kinase inhibitors-associated datasets, ML and DL algorithms such as naive Bayesian (NB) [22–24], k-nearest neighbors (KNN) [24–26], random forest (RF) [27–30], support vector machine (SVM) [25, 26, 31], and deep neural network (DNN) [32, 33] have been used to construct models on the basis of various molecular descriptors and fingerprints for predicting a larger spectrum of kinases inhibition activities for a molecule. These established models play a key role in the theoretical prediction of kinase profiling due to their accuracy and speed of prediction results, and have accelerated the identification and optimization of kinase inhibitors in the early stage of drug discovery.

However, the existing kinase profiling models have the following shortcomings. Firstly, there are two major flaws in the modelling dataset for the kinase profiling prediction task. For one thing, the number of kinases involved in constructing the kinase profiling prediction models is small, limiting its versatility (narrow kinome prediction) compared to the human kinome containing more than 500. For example, the kinase profiling prediction models proposed by Bora and coworkers only includes 107 kinases [29, 34]. For another, the number of compounds in dataset are relatively small, which may lead to the limited generalization ability of the established models. For example, in 2020, Li et al. [34] proposed a virtual kinase profiling model against a panel of 391 kinases, however, there are approximately 40 kinases with less than 10 compounds (actives and inactives). Apparently, the predictive models based on these insufficient compound datasets may not achieve good generalization performance. Secondly, for different tailored modelling datasets, the existing models are constructed based on a specific molecular representation (i.e. molecular descriptors or fingerprints) by using only single or limited ML methods. Obviously, this lack of combined screening of molecular features and ML algorithms will result in the built models that may not be able to achieve the highest accuracy. In other words, it is impossible to assess which ML methods can achieve higher performance in building kinase profiling models from the existing studies. Thirdly, most of the existing kinase profiling predictive models are trained using conventional ML (e.g., KNN, NB, SVM and RF) algorithms, hile the advanced DL (especially graph neural network, GNN) algorithms, which have been successfully used to predict molecular properties and bioactivities, have seldom conducted for the kinase profiling prediction [35–38]. In addition, the reported kinase profiling predictive models have not been integrated into easy-to-use tools (e.g., local software package or online platform), which limits the use of these models by experts and non-experts in the field.

To address the above-mentioned shortcomings regarding the kinase profiling prediction task, herein, we constructed a comprehensive kinase profiling prediction benchmark dataset (called KinaseNet) from multiple sources for 354 kinases. A total of 136,290 predictive models were then built based on three types of molecular representations (i.e. a set of molecular descriptors, five different molecular fingerprints, and molecular graphs) using five mainstream ML methods (e.g., KNN [39], NB [40], SVM [41], RF [42], and XGBoost [43]) and seven advanced DL algorithms including DNN [44], graph convolutional network (GCN) [45], graph attention network (GAT) [46], message passing neural networks (MPNN) [47], Attentive FP [48], D-MPNN (Chemprop) [49] and FP-GNN [50]. The performances of these ML and DL models were comprehensively compared and evaluated.

The influences of the sizes of the modelling datasets and features selection on the performances of the kinase profiling models are also explored. Finally, the best models based on the comprehensive comparison results were used to develop an online platform and its python software for supporting kinase inhibitor drug discovery related tasks. The scheme and workflow of this work are shown in Fig. 1.

Materials and methods

Benchmark dataset for kinase profiling prediction

All quantitative compound-kinase associations were collected from ChEMBL (Version 29) [51], PubChem [52], BindingDB [53], and Zinc [54]. We then processed the raw data using the following steps: (1) only ATP-competitive kinase inhibition assay data (assay type: B) for each compound was kept, and compounds with detailed biological activities recorded as IC_{50}, EC_{50}, K_d, or K_i were maintained; (2) the bioactivity units (g/mL, M, and nM) were translated to the standard unit in μM, molecules whose labels could not be unequivocally assigned (e.g., IC_{50}, EC_{50}, K_i, or $K_d < 100$ or > 1 μM) were excluded; and if a compound has multiple inhibitory activity test data for a kinase, we averaged the reported bioactivity records as the final inhibitory activity value; (3) all molecular structures in the kinase profiling dataset were processed using the Standardizer package (https://github.com/flatkinson/standardiser, version 0.1.9), including removal of counter ions, solvent fractions and salts, and adding hydrogen atoms, and once all molecules

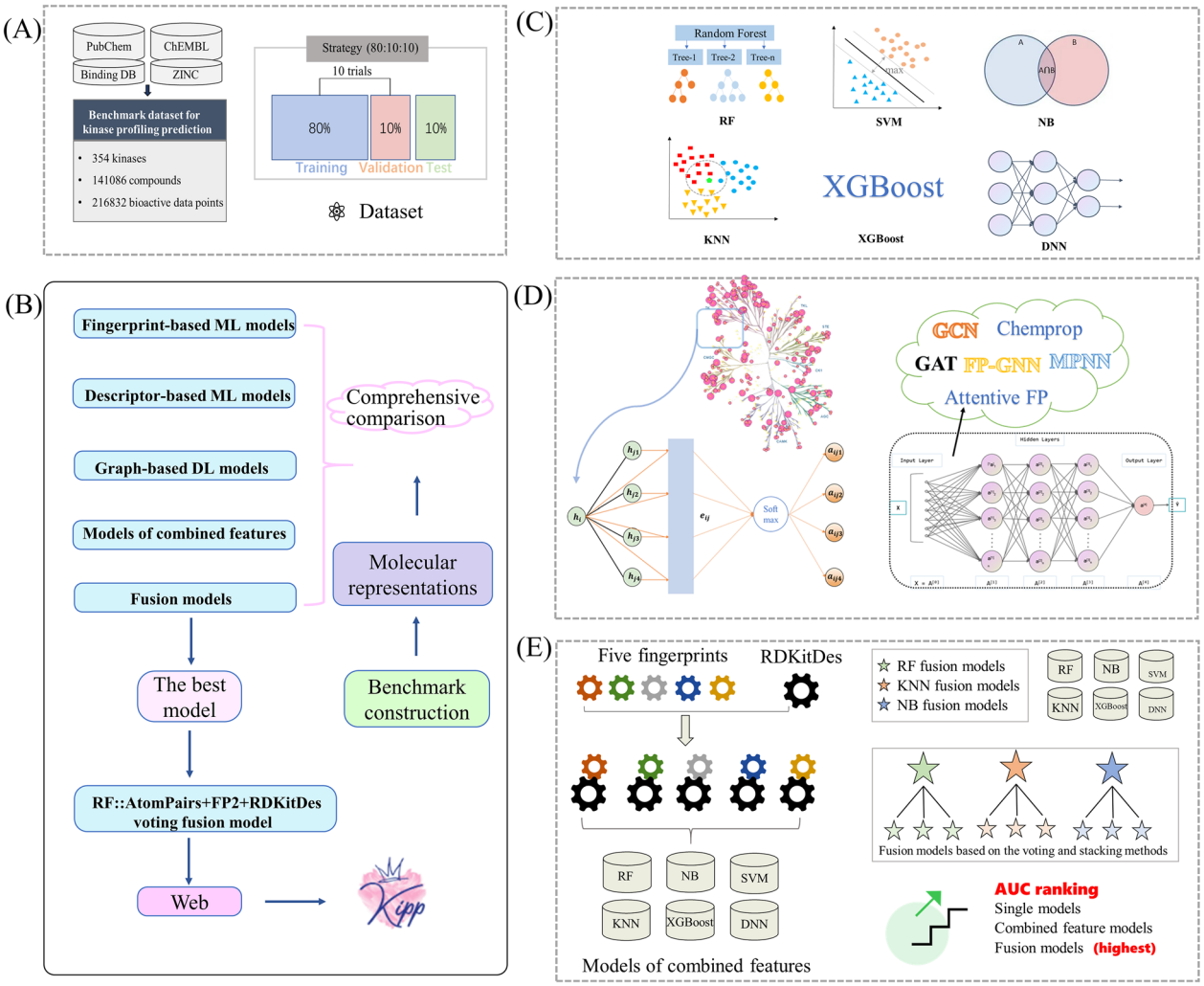

Fig. 1 The scheme and workflow of this work. **A** Dataset collection. **B** Models construction and selecting the optimal model for the kinase profiling prediction task. **C** ML methods for the construction of fingerprint- and RDKitDes-based models. **D** DL methods for the construction of graph-based models. (E) Combined-features- and fusion-based models construction

were standardized, those with molecular weight greater than 1000 Da as well as duplicated molecules were removed; (4) compounds were labeled as actives (pK_i/pK_d/pIC_{50}/$pEC_{50} \geq 6$) and inactives (pK_i/pK_d/pIC_{50}/$pEC_{50} < 6$) in each kinase [34, 55], and we preserved compound − kinase associations only for those kinases with at least 20 active molecules. After applying those criteria, the final comprehensive kinase profiling dataset consists of 141,086 molecules with 216,823 bioactive data points for 354 kinases. Each kinase dataset was randomly divided into three sub-datasets: training set (80%), validation set (10%), and test set (10%). The modelling datasets utilized in the present study are freely available at https://kipp.idruglab.cn/about.

Molecular representations calculation

In this study, five molecular fingerprints including Morgan fingerprints (ECFP-like, 1024-bits) [56], MACCS keys (166-bits) [57], AtomParis fingerprints (1024-bits) [58], FP2 fingerprints (1024-bits) [59] and 2D pharmacophore fingerprints (PharmacoPFP, 38-bits) [60] were used to construct fingerprint-based predictive models. A set of 208 RDKit molecular descriptors (termed RDKitDes) was chosen for the development of descriptor-based predictive models. The fingerprints and descriptors were calculated using open source RDKit software (http://www.rdkit.org/, version: 2020.03.1).

In a molecular graph, the atomic and atomic pair features are used together as a feature matrix [61]. Chemprop and FP-GNN utilize RDKit software (version: 2020.09.5) to calculate molecular graphs. Other molecular graph-based representations were generated using DeepChem (version: 2.5.0). For example, the MolGraphConvFeatureizer module was used to calculate the molecular graphs for the GAT, MPNN, and Attentive FP models, while the ConvMolFeaturizer [62] module was used to compute the molecular graph representation for GCN models.

Selection of ML and DL algorithms for the assessment and model construction

Five mainstream ML and seven advanced DL algorithms were used to build the kinase profiling predictive modes for 354 kinases. These modelling methods (Table 1) are briefly introduced as follows.

Table 1 Detailed ML and DL modelling methods used in this study

Method	Molecular feature	Hyperparameter optimization	Website
RF[a]	RDKitDES or fingerprints (Morgan, MACCS, AtomPairs, FP2, and PharmacoPFP)	Grid search	https://github.com/scikit-learn/scikit-learn
NB[b]		Grid search	https://github.com/scikit-learn/scikit-learn
SVM[c]		Grid search	https://github.com/scikit-learn/scikit-learn
KNN[d]		Grid search	https://github.com/scikit-learn/scikit-learn
XGBoost[e]		Grid search	https://github.com/dmlc/xgboost
DNN[f]		Grid search	https://deepchem.io/
GCN[g]	molecular graphs	Grid search	https://deepchem.io/
GAT[h]	molecular graphs	Grid search	https://deepchem.io/
MPNN[i]	molecular graphs	Grid search	https://deepchem.io/
Attentive FP[j]	molecular graphs	Grid search	https://deepchem.io/
Chemprop[k]	molecular graphs	Bayesian Optimization	https://github.com/chemprop/chemprop
FP-GNN[l]	molecular graphs and fixed molecular fingerprints (MACCS, PubChem, and Pharmacophore ErG fingerprints)	Bayesian optimization	https://github.com/idrugLab/FP-GNN

[a] RF: Random forest

[b] NB: Naïve Bayesian

[c] SVM: Support vector machine

[d] KNN: K-Nearest Neighbor

[e] XGBoost: Extreme gradient boosting

[f] DNN: Deep neural networks

[g] GCN: Graph convolutional network

[h] GAT: Graph attention network

[i] MPNN: Message passing neural networks

[j] Attentive FP

[k] Chemprop: D-MPN

[l] FP-GNN

Random forest (RF)

RF, developed by Svetnik et al.[42], is an ensemble recursive partitioning approach in which each recursive partitioning 'tree' is built from a bootstrapped sample of compounds, and each branch of a tree uses a random subset of descriptors [27]. The following five hyperparameters were tuned to achieve the optimal RF model: n_estimators (10–500), criterion ('*gini*' and '*entropy*'), max_depth (0–15), min_samples_leaf (1–10), and max_features ('*log2*', '*auto*' and '*sqrt*').

Naïve Bayesian (NB)

NB classifier is developed based on Bayes' theorem [40] and widely used in molecular properties prediction and virtual screening (VS) projects [63–66]. Two hyperparameters were optimized for NB models construction: alpha (0.01–1) and binarize (0, 0.5, 0.8).

Support vector machine (SVM)

SVM was formally developed in 1995 [41] and quickly became a mainstream ML method due to its excellent performance in text classification tasks [67]. The principle of SVM is to determine the optimal hyperplane in the feature space by maximizing the boundaries between classes in *N*-dimensional space, which can distinguish objects with various class labels. Two hyperparmeters, Kernel coefficient (gamma, 'auto', 0.1–0.2) and penalty parameter C of the error term (C, from 1 to 100), were optimized for the development of SVM models.

K-nearest neighbor (KNN)

KNN is a commonly used supervised learning method with a simple mechanism. For a given test sample, it finds the *k* closest training samples in the training set based on distance measures (e.g., Manhattan, Euclidean, and Jaccard distance), and then makes a prediction based on the information of these *k* 'neighbors' [39]. In the training of KNN models, the default Euclidean distance metric was utilized, and three hyperparameters including n_neighbors (1–5), p (1–2), and weight function ('uniform', 'distance'), were optimized.

Extreme gradient boosting (XGBoost)

XGBoost is one of the most representative ensemble ML algorithms under the gradient boosting framework [43]. It has been shown to achieve state-of-the-art (SOTA) performance on many standard classification benchmark datasets [37, 68, 69]. Seven hyperparameters were optimized: learning_rate (0.01–0.1), n_estimators (50–100), max_depth (3–5), min_child_weight (1–3), gamma (0–0.1), subsample (0.8–1.0), and colsample bytree (0.8–1.0).

Deep neural networks (DNN)

DNN is essentially an artificial neural network with an input layer, an output layer, and multiple hidden layers, which mimics the behavior of biological neural networks [44]. DNN consists of a large number of individual neurons [70, 71], and each neuron in the DNN architecture collects information from its associated neurons and a non-linear activation function was then used to activate the aggregated information. Three hyperparameters were optimized: dropouts (0.1, 0.2, 0.5), layer_sizes (64, 128, 256, 512) and weight_decay_penalty (0.01, 0.001, 0.0001).

Graph convolutional network (GCN)

GCN uses graph-structured data as features input [45], and consists of graph convolution layers, a readout layer, fully linked layers, and an output layer. The basic principle of GCN is to use edge information to aggregate node information, resulting in a new node representation. Several frameworks of GCN and variants have been proposed so far. For example, Duvenaud et al. [62] proposed a convolutional neural network that operates directly on molecular graphs, allowing end-to-end learning of prediction pipelines to exhibit better predictive performance for molecular property prediction tasks. Here, this GCN architecture was used to establish GCN models, and the following hyperparameters were optimized: weight decay (0, 10e-8, 10e-6, 10e-4), graph conv layers ([64, 64], [128, 128], [256, 256], learning rate (0.01, 0.001, 0.0001), and dense layer size (64, 128, 256).

Graph attention network (GAT)

GAT introduces an attention mechanism based on the GCN [46], which calculates the weights of the features of nodes and adjacent nodes through aggregation, and follows a self-aggregation strategy. GAT can better extract the spatial feature relationships of nodes compared to the GCN in the application of directed graphs [72]. Six hyperparameters were optimized in the training of the GAT models, including weight_decay (0, 10e-8, 10e-6, 10e-4), learning rate (0.01, 0.001, 0.0001), n_attention_heads (8, 16, 32), and dropouts (0, 0.1, 0.3, 0.5).

Message passing neural network (MPNN)

MPNN, first proposed by Gilmer and coworkers in 2017 [47], represents a commonly used GNN framework for various chemical prediction tasks. Many new GNN architectures have been developed based on the excellent performance and flexibility of MPNN framework for molecular property prediction [49, 73–75]. Herein, the main hyperparameters were optimized as follows: weight_decay (10e-8, 10e-6, 10e-4), learning rate (0.01, 0.001, 0.0001), graph_conv_layers ([64, 64], [128, 128],

[256, 256]), num_layer_set2set (2, 3, 4), node_out_feats (16, 32, 64), and edge_hidden_feats (16, 32, 64).

Attentive FP

Attentive FP is an advanced GNN model that allows the model to focus on the most important elements of the input using graph attention mechanism [48]. It has been reported to exhibit SOTA performance for predicting molecular properties. Herein, the primary hyperparameters including dropout (0, 0.1, 0.5), graph feat size (50, 100, 200), num timesteps (1, 2, 3), num layers (2, 3, 4), learning rate (0.0001, 0.001, 0.01), and weight decay (0, 0.01, 0.0001), were optimized for the development of the Attentive FP models.

D-MPNN (Chemprop)

D-MPNN (Chemprop) was developed upon the MPNN framework by adopting a message-passing paradigm based on updating representations of directed bonds rather than atoms [49]. Chemprop has been successfully applied for the discovery of structurally distinct antibiotics [76]. Herein, the hyperparameters were optimized as follows: dropout (2, 3), dropout gat (0, 0.05), dim (1, 2), and gat scale (300, 400).

FP-GNN

Recently, FP-GNN as a novel DL architecture [50] was developed in our Lab for enhanced molecular properties prediction. FP-GNN not only learns to characterize the local atomic environment by propagating node information from nearby nodes to more distant nodes using the attention mechanism in a task-specific encoding, but also simultaneously learns a strong prior knowledge based on the fixed and complementary molecular fingerprints (MACCS, PubChem, and Pharmacophore ErG fingerprints). We used FP-GNN algorithm to build models for the kinase profiling prediction task. The hyperparameters were optimized as the following: dropout (0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6), dropout gat (0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6), dim (300, 350, 400, 450, 500, 550, 600), gat scale (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8), nheads (2, 3, 4, 5, 6, 7, 8), and nhid (40, 45, 50, 55, 60, 65, 70, 75, 80).

The RF, SVM, KNN, and NB models were constructed using the Scikit-learn python package (https://github.com/scikit-learn/scikit-learn, version: 0.24.1) [77]; the XGBoost models were developed using the XGBoost python package (https://github.com/dmlc/xgboost, version: 1.3.3) [43]; four graph-based models (GCN, GAT, MPNN and Attentive FP) were established using the DeepChem python package (https://deepchem.io/); D-MPNN (Chemprop) models were constructed using the Chemprop python package (https://github.com/chemprop/chemprop); and FP-GNN models were developed using the FP-GNN software (https://github.com/idrugLab/FP-GNN). All ML and DL models were trained on CPU (Intel(R) Xeon(R) Silver 4216 CPU@2.10 GHz) and GPU (NVIDIA Corporation GV100GL [Tesla V100 PCIe 32 GB]), respectively. Additionally, Bayesian optimization was applied to optimize hyperparameters for FP-GNN and Chemprop models, while grid search method was employed to optimize hyperparameters for other models.

Performance evaluation metric

To benchmark the performance of different ML and DL tools for the kinase profiling prediction, six metrics, including specificity (SP/TNR), sensitivity (SE/TPR/Recall), Balanced accuracy (BA), F1 score, Matthew's correlation coefficient (MCC), and area under the receiver operating characteristic (ROC) curve (AUC), are used and defined as follows:

$$\mathrm{SP} = \frac{TN}{TN + FP} \tag{1}$$

$$\mathrm{SE} = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{BA} = \frac{TPR + TNR}{2} = \frac{SE + SP}{2} \tag{3}$$

$$\mathrm{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{4}$$

$$\mathrm{MCC} = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN) \times (TP + FP) \times (TN + FN) \times (TN + FP)}} \tag{5}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

AUC was the most commonly used criterion for kinase inhibitor activity prediction tasks [15, 29, 30, 34, 35, 78], we therefore selected AUC value as the indicator of the accuracy of the classification models for a fair comparison. Given that active compounds outnumbered inactive compounds in the current kinase profiling modelling dataset, with a positive-to-negative ratio of 3.83, F1 score was also utilized to judge the accuracy of the models [34, 79–81].

Results and discussion

Benchmark dataset analysis and model construction

We obtained a comprehensive kinase profiling modelling dataset from multiple sources by applying the criteria

mentioned in the Methods section. This dataset contains 141,086 unique molecules involving in 216,823 inhibitory activity data points, which covers 354 kinases from nine groups in the human kinome: TK family (88 kinases), CMGC family (48 kinases), AGC family (44 kinases), CAMK family (46 kinases), STE family (38 kinases), TKL family (30 kinases), Atypical family (16 kinases), CK1 family (6 kinases), and Others (38 kinases), Detailed information of the dataset are shown in Additional file 2: Table S1. The average ratio of positive (actives) to negative (inactives) was approximately 3.83, implying that the modelling dataset is relatively unbalanced. Nonetheless, in order to objectively explore and evaluate the predictive performance of different computational methods, we preferred to utilize the raw data from experimentally validated molecules against these kinases, without adding theoretical decoys to deliberately balance the modelling dataset. Bemis–Murcko scaffold analysis was conducted to analyze the structural diversity of molecules in the dataset. The proportion of scaffolds to molecules for each kinase falls between 10 and 100%, with an average value of 51.0%, suggesting that the molecules of the dataset were structurally diverse. Besides, compounds have broad distributions of molecular weight (36.461–998.013) and AlogP (-8.895–11.509), indicating that the compounds in the modelling dataset have an extensive chemical space (Additional file 2: Table S2). Such results imply that the predictive models based on this dataset could exhibit better reliability and robustness.

For this comprehensive kinase profiling modelling dataset, a total of 148,680 classification predictive models were generated based on the three different types of molecular features using the selected 12 ML and DL algorithms. To fairly compare the performance of the ML and DL methods for the kinase profiling predictive task, the average of the evaluation metrics of the established models for each algorithm were calculated as the final result. The details of performance of the established models are described and discussed in the following sections.

Performance evaluation results of fingerprint-based ML and DL models

Five ML (KNN, NB, RF, SVM, and XGBoost) and one DL (DNN) approaches were used to build 106,200 predictive models based on five types of fingerprints (Morgan, MACCS, AtomPairs, FP2 and PharmacoPFP). Each model is denoted as a combination of the ML method and the corresponding molecular representation (e.g., DNN::Morgan).

As shown in Table 2, most of the fingerprint-based models performed well for the kinase profiling predictive task, with an average AUC value > 0.73 and average F1 value > 0.72 on the test sets. Despite the differences in the characteristics of the five molecular fingerprints, the RF method performed the best for 354 kinases (Fig. 2), with the highest average AUC value (0.769) and MCC value

Table 2 Performance comparison results of the fingerprint-based models on the test sets of 354 kinases

Molecular feature	Method	AUC [g]	F1 score [h]	BA [i]
AtomPairs	RF[a]	0.779 ± 0.161	0.736 ± 0.259	0.625 ± 0.124
	NB[b]	0.733 ± 0.135	0.716 ± 0.186	0.680 ± 0.117
	SVM[c]	0.698 ± 0.214	0.712 ± 0.286	0.620 ± 0.157
	KNN[d]	0.743 ± 0.152	0.747 ± 0.222	0.665 ± 0.126
	XGBoost[e]	0.759 ± 0.167	0.750 ± 0.212	0.653 ± 0.127
	DNN[f]	0.752 ± 0.171	0.714 ± 0.238	0.631 ± 0.128
	Mean	0.744 ± 0.027	0.729 ± 0.017	0.646 ± 0.024
FP2	RF	0.786 ± 0.150	0.731 ± 0.258	0.634 ± 0.118
	NB	0.743 ± 0.141	0.728 ± 0.173	0.692 ± 0.121
	SVM	0.682 ± 0.259	0.686 ± 0.288	0.590 ± 0.191
	KNN	0.748 ± 0.149	0.760 ± 0.200	0.671 ± 0.121
	XGBoost	0.761 ± 0.163	0.752 ± 0.218	0.659 ± 0.125
	DNN	0.753 ± 0.179	0.722 ± 0.237	0.626 ± 0.132
	Mean	0.746 ± 0.035	0.730 ± 0.026	0.645 ± 0.036
MACCS	RF	0.751 ± 0.166	0.732 ± 0.257	0.613 ± 0.121
	NB	0.724 ± 0.142	0.720 ± 0.177	0.662 ± 0.117
	SVM	0.670 ± 0.253	0.681 ± 0.292	0.577 ± 0.190
	KNN	0.719 ± 0.147	0.750 ± 0.201	0.646 ± 0.119
	XGBoost	0.739 ± 0.168	0.741 ± 0.224	0.639 ± 0.124
	DNN	0.705 ± 0.181	0.697 ± 0.249	0.591 ± 0.121
	Mean	0.718 ± 0.028	0.720 ± 0.027	0.621 ± 0.033
Morgan	RF	0.774 ± 0.166	0.722 ± 0.282	0.612 ± 0.122
	NB	0.772 ± 0.143	0.745 ± 0.176	0.702 ± 0.124
	SVM	0.680 ± 0.268	0.685 ± 0.292	0.594 ± 0.192
	KNN	0.755 ± 0.154	0.755 ± 0.211	0.674 ± 0.124
	XGBoost	0.761 ± 0.164	0.749 ± 0.223	0.653 ± 0.128
	DNN	0.761 ± 0.176	0.715 ± 0.245	0.621 ± 0.132
	Mean	0.751 ± 0.035	0.729 ± 0.027	0.643 ± 0.041
PharmacoPFP	RF	0.757 ± 0.174	0.735 ± 0.258	0.620 ± 0.121
	NB	0.726 ± 0.144	0.722 ± 0.174	0.670 ± 0.123
	SVM	0.684 ± 0.240	0.689 ± 0.281	0.587 ± 0.184
	KNN	0.740 ± 0.147	0.761 ± 0.193	0.664 ± 0.120
	XGBoost	0.748 ± 0.175	0.745 ± 0.225	0.649 ± 0.129
	DNN	0.735 ± 0.183	0.709 ± 0.249	0.614 ± 0.130
	Mean	0.732 ± 0.026	0.727 ± 0.026	0.634 ± 0.032

[a] RF: Random forest

[b] NB: Naïve Bayesian

[c] SVM: Support vector machine

[d] KNN: K-Nearest Neighbor

[e] XGBoost: Extreme gradient boosting

[f] DNN: Deep neural networks

[g] AUC: Area under the receiver operating characteristics curve

[h] F1 scores: F1-measure

[i] BA: Balanced accuracy. "±" values represent standard deviations

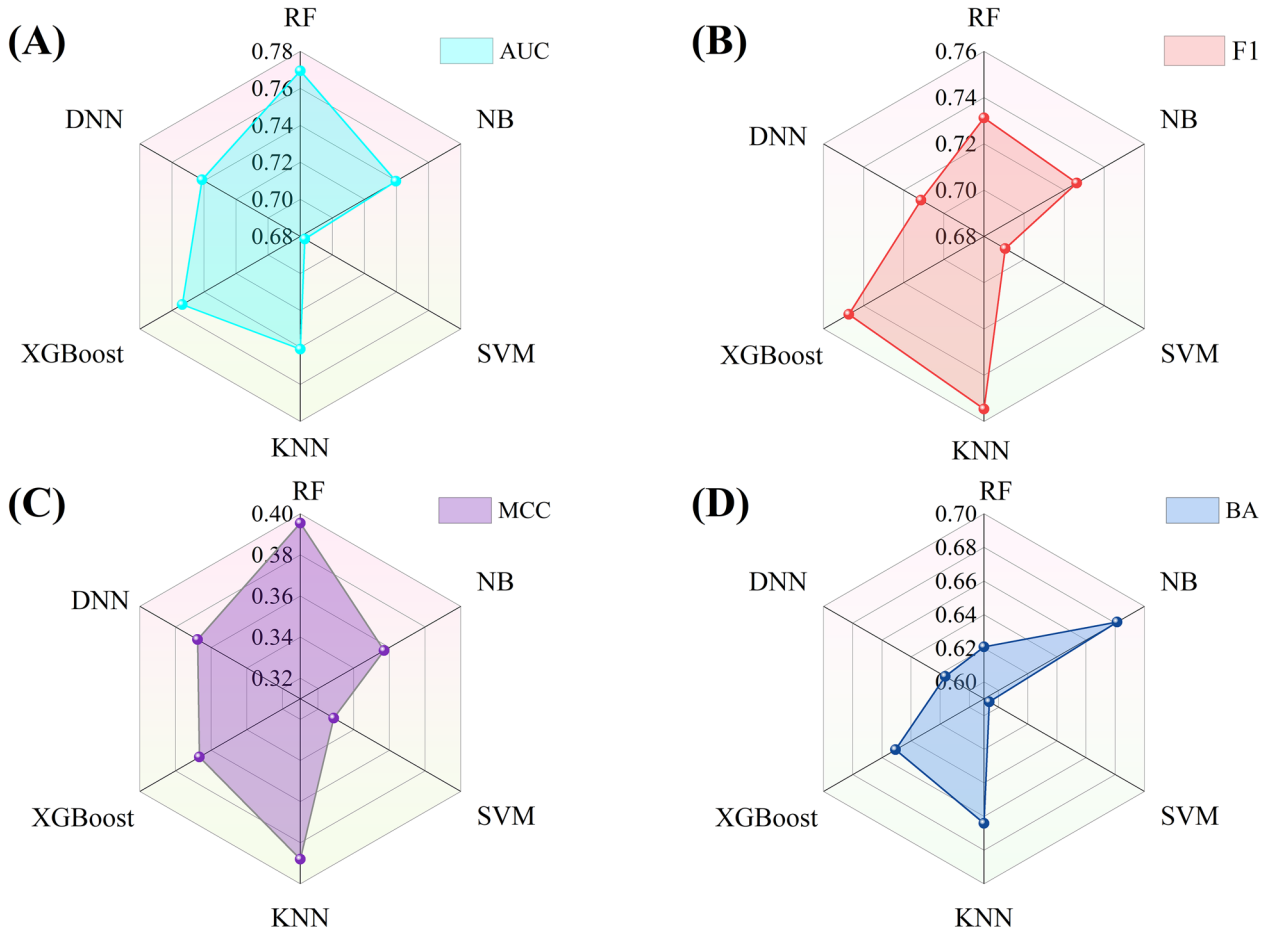

Fig. 2 Performance comparison results of fingerprint-based models using different ML algorithms. **A**, **B**, **C** and **D** represent the comparison results based on the average F1 score, AUC, BA, and MCC values from the test sets, respectively

(0.395), and relatively high F1 score (0.731) and BA value (0.621). In addition, another ensemble learning methods, XGBoost, also showed considerable predictive performance, achieving the second highest AUC value (0.754) and F1 score (0.747), and relatively high BA value (0.651) and MCC value (0.367).

The Morgan fingerprints achieved highest mean AUC value (0.751 ± 0.035, Table 2), which implies that it is a relatively better molecular representation for kinase profiling prediction. In addition, combining different ML methods and different molecular fingerprints yielded different performance results, indicating that it is necessary to screen the combination of modelling algorithms and feature expressions to achieve the best performance. For example, the RF and XGBoost algorithm tends to use the FP2 fingerprints as input features to achieve the best model rather than the Morgan fingerprints. In contrast, the NB algorithm tends to utilize the Morgan fingerprints as input features to generate the best models rather than the FP2 fingerprints (Table 2).

We further analyzed the interval distribution of the average AUC values of the test sets of 354 kinase targets for each method. As shown in Fig. 3, although different combinations of fingerprints and modelling methods can produce different distributions of AUC values, statistical analysis found that the AUC values of the majority of the fingerprint-based models (~72.2%) were greater than 0.7. For example, the numbers of high quality (HQ, AUC > 0.7) for the RF::AtomPairs and XGBoost::AtomPairs models were 262 (Fig. 3A) and 248 (Fig. 3E) kinases, respectively. In addition, the RF::FP2 models showed obvious advantage, achieving the highest average AUC value (0.786 ± 0.150, Table 2). Importantly, it can achieve AUC values greater than 0.7 on 269 kinases (Fig. 3A).

The Morgan fingerprints owns the relatively better predictive performance with highest average AUC value, however, this does not necessarily mean that other fingerprints cannot outperform the Morgan fingerprints on individual kinases. Figure 4A showed that the FP2,

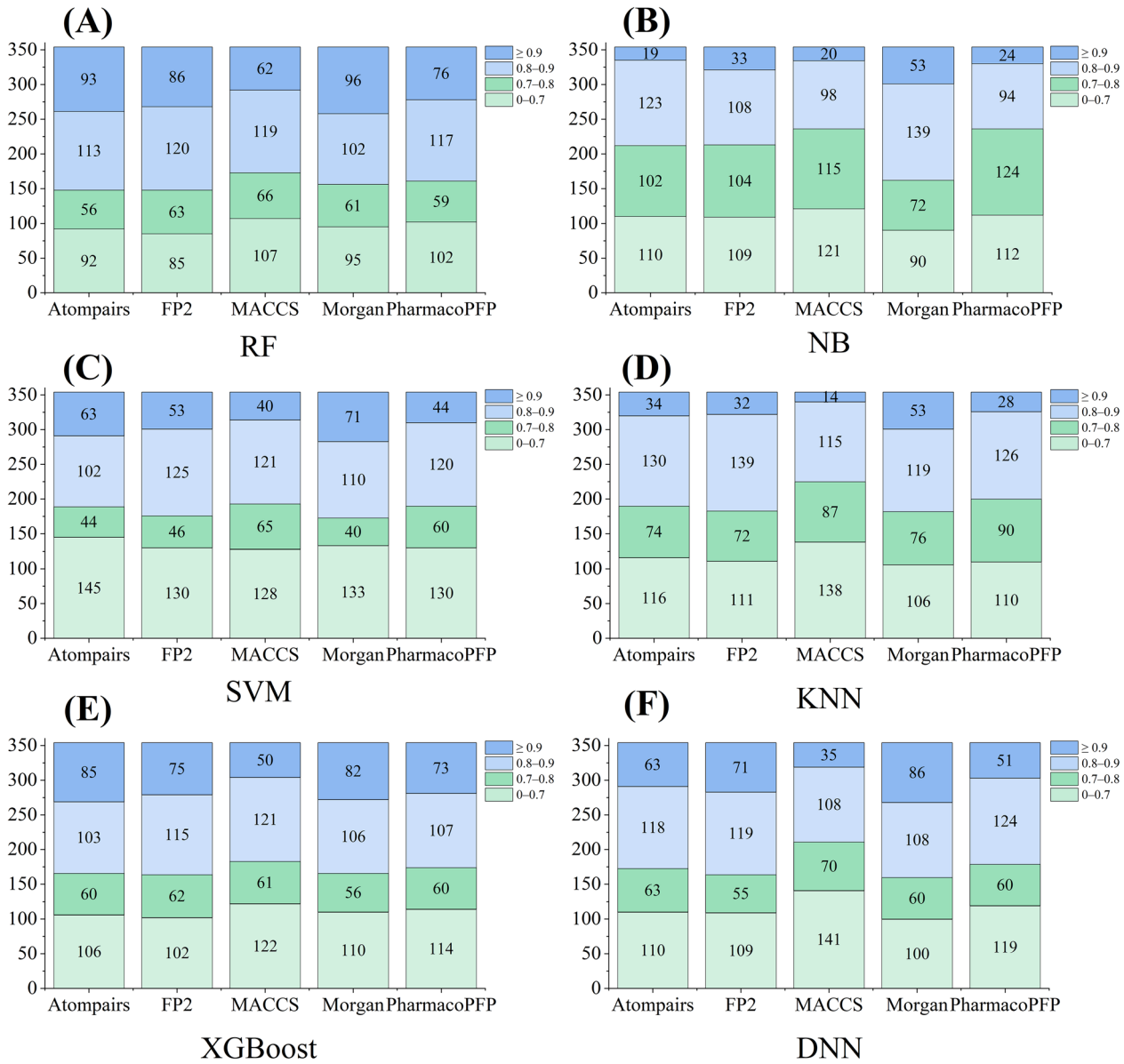

Fig. 3 The interval distribution of the AUC values of fingerprint-based models for 354 kinases by using RF (**A**), NB (**B**), SVM (**C**), KNN (**D**), XGBoost (**E**), and DNN (**F**) algorithms

AtomPairs, MACCS, and PharmacoPFP fingerprints contributed eight, eight, two, and two unique kinase targets in the models with an AUC ≥ 0.8. Although the Morgan fingerprints also contributed the most models with an AUC ≥ 0.8, and the majority of these models were commonly found by at least two of other four fingerprints (i.e. FP2, MACCS, Morgan and PharmacoPFP fingerprints). The most unique HQ models was obtained by the AtomPairs fingerprints with an average AUC greater than 0.9 (Fig. 4B), i.e. the FP2, MACCS, Morgan and PharmacoPFP fingerprints can generates two, three, six, and seven unique HQ models that cannot be obtained by the AtomPairs fingerprints.

Recently, Merget et al. [30] reported RF models based Morgan fingerprints for the profiling prediction of kinase inhibitors, with an average AUC of 0.76 on 291 kinases, and achieving HQ (AUC > 0.7) on ~ 200 kinases. Apparently, the RF::FP2 models proposed in this study are superior to the models from Merget et al. study in terms of the total of number of kinases (354) and the overall accuracy (mean AUC = 0.786), as well as the number of HQ models (269, AUC > 0.7). In addition, the RF::Morgan

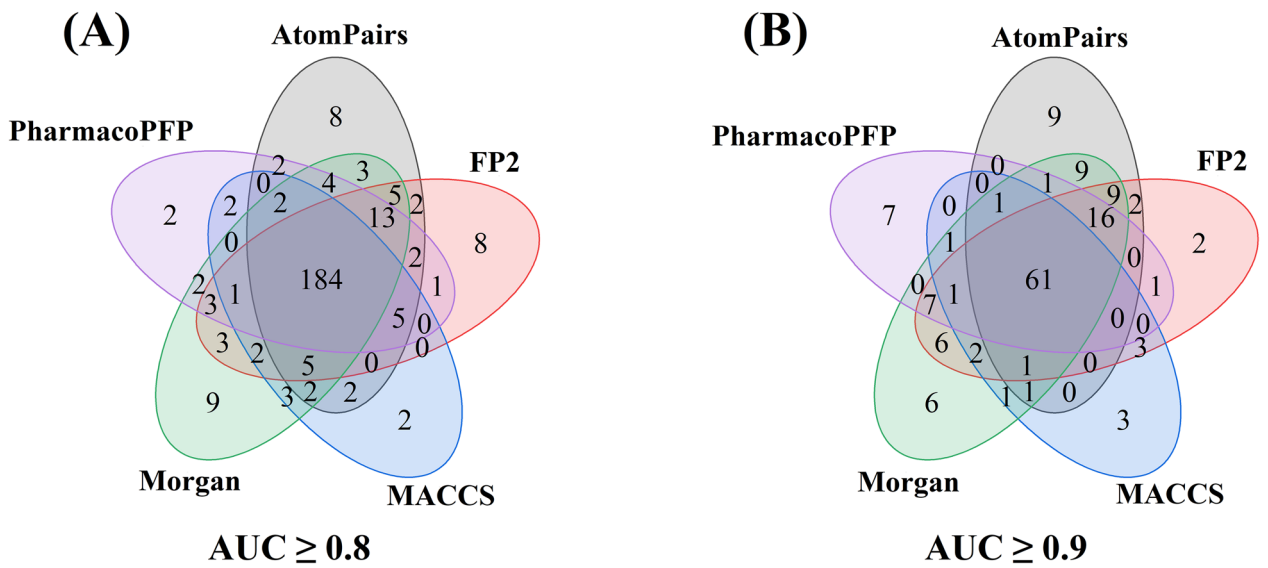

Fig. 4 Overlap analyses of various fingerprint-based high-quality (HQ) models with an average AUC of ≥ 0.8 (**A**) and ≥ 0.9 (**B**), respectively

models proposed herein have comparable or superior performance to the models of Merget et al., i.e. it exhibited average AUC value of 0.774 on 354 kinases and achieved HQ models on 259 kinases. The results illustrated that the comprehensive kinase profiling dataset with large structural diversity and chemical space constructed in this paper is necessary for building robust and reliable kinase profiling prediction models, as well as the optimal combination of ML algorithms and molecular feature representations can help to develop more accurate models for the virtual profiling prediction of kinase inhibitors.

Performance evaluation results of descriptor-based ML and DL models

Subsequently, a total of 21,240 descriptor-based predictive models were successfully constructed and compared using the same modelling methods. The optimized RDKit-descriptors obtained using the SelectPercentile module (Percentile = 30) implemented in the scikit-learn package were utilized as input features for model construction. Detailed performance results of the descriptor-based models are listed in Additional file 2: Table S3. The average F1, AUC, and BA values for the test sets of these models are summarized in Table 3.

As shown Table 3, most descriptor-based predictive models performed quite well, with mean F1 scores = 0.74, and average AUC value greater than 0.75. In accordance with the fingerprint-based models evaluation results where RF method achieved the best performance, RF::RDKitDes also performed best with the highest average AUC value (0.798 ± 0.120) (Table 3) on these

Table 3 Performance comparison results of RDKit descriptor-based predictive models on the test sets of 354 kinases

Molecular feature	Method	AUC[g]	F1 score[h]	BA[i]
RDKitDes	RF[a]	0.798 ± 0.120	0.759 ± 0.225	0.650 ± 0.113
	NB[b]	0.763 ± 0.099	0.739 ± 0.155	0.681 ± 0.090
	SVM[c]	0.727 ± 0.206	0.723 ± 0.245	0.611 ± 0.165
	KNN[d]	0.774 ± 0.116	0.776 ± 0.186	0.684 ± 0.104
	XGBoost[e]	0.755 ± 0.148	0.747 ± 0.216	0.650 ± 0.117
	DNN[f]	0.718 ± 0.180	0.693 ± 0.254	0.589 ± 0.117
	Mean	0.756 ± 0.030	0.740 ± 0.029	0.644 ± 0.038

[a] RF: Random forest

[b] NB: Naïve Bayesian

[c] SVM: Support vector machine

[d] KNN: K-Nearest Neighbor

[e] XGBoost: Extreme gradient boosting

[f] DNN: Deep neural networks

[g] AUC: Area under the receiver operating characteristics curve

[h] F1 scores: F1-measure

[i] BA: Balanced accuracy. "±" values represent standard deviations

descriptor-based models, which by the way is higher than any other fingerprint-based models (Table 2). According to the average AUC values of these descriptor-based models (Table 3), KNN method achieved the second-ranked predictive performance, followed by NB and XGBoost methods.

Figure 5A illustrates that approximately 73% of the descriptor-based models are HQ models, which outperform the aforementioned fingerprint-based models. Taking the RF::RDKitDes model as an example, it not only

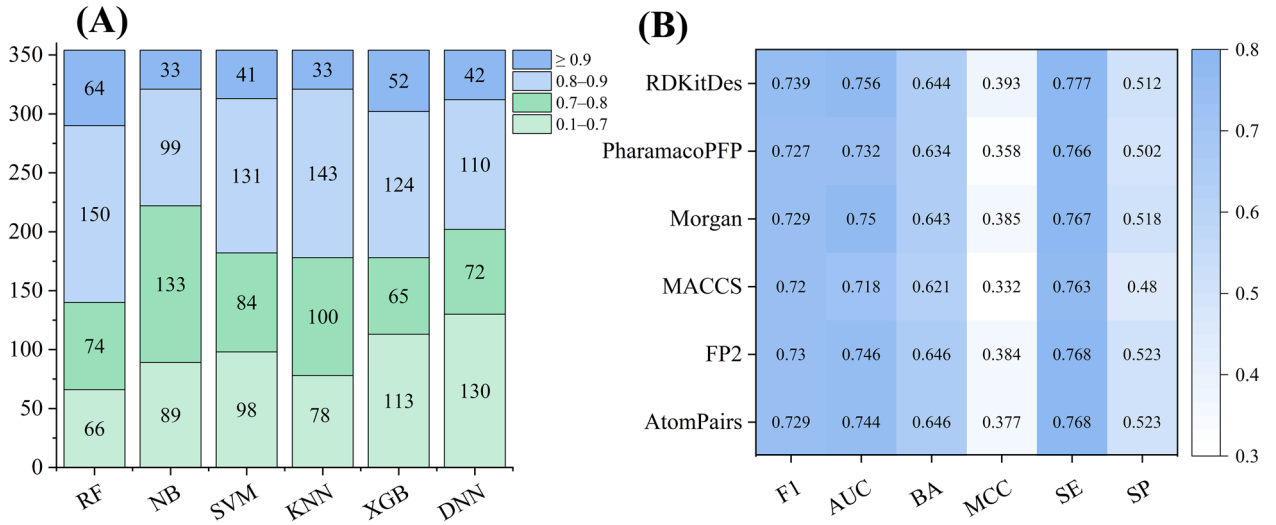

Fig. 5 **A** Detailed distribution of the average AUC values of RDKitDes-based models for 354 kinases. **B** Heatmap analysis results of the average metrics of RDKitDes- and fingerprint-based models on the test sets

achieved the highest mean AUC value, but achieved 288 HQ models (Fig. 5A) for 354 kinases. Clearly, the RF::RDKitDes model outperforms the corresponding RF-based fingerprint models in terms of both the average AUC metric and the number of HQ models (Table 2 and Fig. 3A), regardless of which molecular fingerprints is used as input features.

To further confirm whether descriptor-based models outperform fingerprint-based models, we systematically compare the evaluation metrics of these models. As shown in Fig. 5B, RDKitDes-based models slightly outperformed fingerprint-based models due to their best performances in terms of the high average F1 score, AUC, SE and MCC values. The detailed comparison results of descriptor- and fingerprint-based models for each ML algorithm are shown in Additional file 1: Fig. S1. For example, RDKitDes-based models achieved the highest F1 scores and AUC values on the RF, SVM, and KNN algorithms (Additional file 1: Figs. S1A, C and D), and slightly weaker and/or comparable performance on the NB, XGBoost and DNN methods (Additional file 1: Figs. S1B, E and F), when compared to fingerprint models based on these ML algorithms. These results highlighted that RDKitDes may be suitable for achieving the optimal performance of ML methods in the kinase profiling prediction task.

Performance evaluation results of graph-based DL models

Currently, various graph-based DL algorithms, which have recently been developed and achieved the SOTA performance in molecular property prediction tasks [48, 49, 82], have not been used for the kinase profiling prediction task. Accordingly, we introduced six GNN-based DL algorithms (Table 4) to model the kinase profiling prediction task. As shown in Table 4, GCN exhibited the overall best performance on the test sets compared to other GNN-based DL methods, achieving the highest average AUC (0.729 ± 0.206) and BA (0.604 ± 0.127) values, and second high F1 score (0.658 ± 0.271). A violin plot analysis of the overall AUC values also demonstrated that GCN performed the best (Fig. 6A), followed by FP-GNN and GAT methods.

Further analysis of the distribution of AUC values shows that the GCN models and FP-GNN models exhibited comparable performance in terms of HQ models, achieving 78 models in the interval where the AUC value is greater than 0.9 (Fig. 6B). Additionally, the GCN models and FP-GNN models, respectively, outperformed the RF::RDKitDes models on 140 and 143 kinases in terms of AUC metric (Additional file 2: Tables S4-S5. Consequently, the predictive models based on the GCN and FP-GNN algorithms are more applicable overall compared to other graph-based DL methods.

However, the use of graph-based DL methods (Table 4) may not be suitable as they do not show any advantage in the kinase profiling prediction task compared to the models based on the fixed prior molecular features such as molecular fingerprints (Table 2) and descriptors (Table 3). Even GCN and FP-GNN models only achieved 226 and 203 HQ models ($AUC > 0.7$) for 354 kinase targets. Typically, graph-based DL algorithms have an inherent self-learning mechanism, which may result in poor performance due to the insufficient modelling datasets in individual kinases. To confirm this point, we further

Table 4 Performance comparison results of different graphs-based DL models on the test sets

Molecular feature	Method	AUC[g]	F1 score[h]	BA[i]
Molecular graphs	GCN[a]	0.729 ± 0.206	0.658 ± 0.271	0.604 ± 0.127
	GAT[b]	0.675 ± 0.225	0.636 ± 0.272	0.582 ± 0.145
	MPNN[c]	0.658 ± 0.202	0.621 ± 0.298	0.557 ± 0.128
	Attentive FP[d]	0.674 ± 0.207	0.661 ± 0.295	0.581 ± 0.116
	Chemprop[e]	0.717 ± 0.173	0.640 ± 0.291	0.573 ± 0.108
	FP-GNN[f]	0.704 ± 0.223	0.627 ± 0.367	0.604 ± 0.142
	Mean	0.693 ± 0.028	0.641 ± 0.016	0.584 ± 0.018

[a] GCN: Graph convolutional network

[b] GAT: Graph attention network

[c] MPNN: Message passing neural networks

[d] Attentive FP

[e] Chemprop: D-MPNN

[f] FP-GNN

[g] AUC: Area under the receiver operating characteristics curve

[h] F1 scores: F1-measure

[i] BA: Balanced accuracy. "±" values represent standard deviations

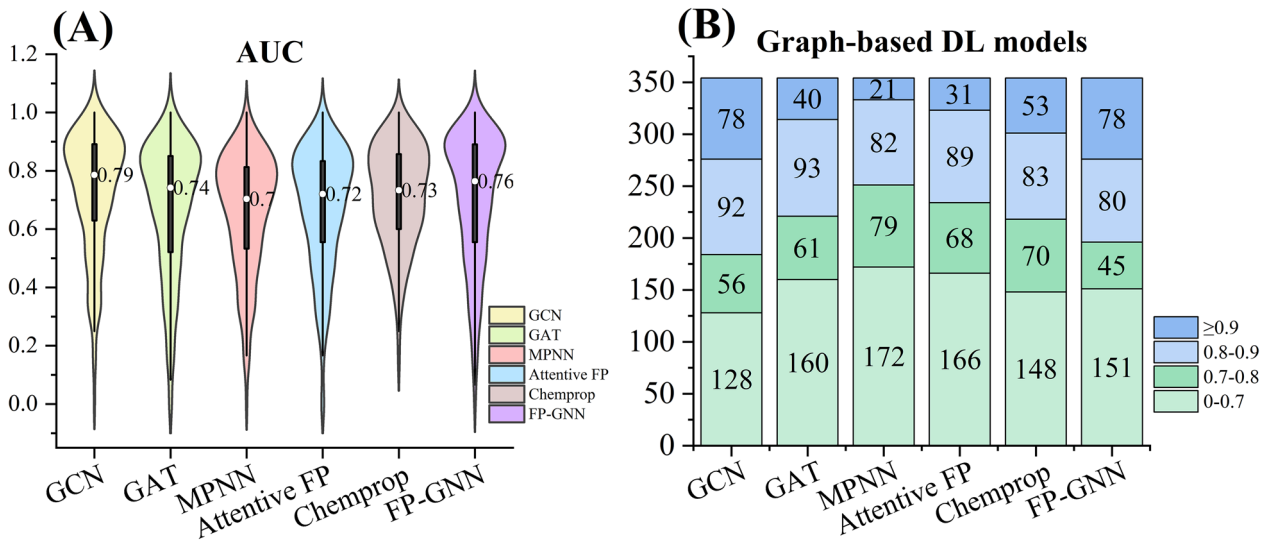

Fig. 6 **A** Violin plot of the overall distribution of AUC values for six graph-based DL models. White spheres represent the medians, and boxes represents 1.5 the interquartile range (1.5IQR) from the median. **B** Detailed distribution of the average AUC values of different graph-based DL models on 354 kinases

analyze whether the size of the modelling dataset for each kinase has an impact on the accuracy of the graph-based DL models. Figure 7 summarizes the relationship between the AUC values in the test sets and compound quantity intervals in the training sets for the graph-based DL models. In general, the prediction performance is positively correlated with the number of compounds in the training set. Taking the GCN method as an example (Fig. 7A), if the number of molecules in modelling dataset is less than 100, few HQ models can be obtained. Similar phenomena are observed in other DL methods (Figs. 7B–F), albeit with some differences. In other words, graph-based DL models possibly acquire better predictive performance on large datasets. Our findings further illustrate the shortcomings of graph-based DL algorithms in the field of kinase prediction, especially for kinases with insufficient activity data. In the future, as the number of kinases and their inhibitors continues to increase, graph-based DL algorithms may be more suitable for many individual kinases to achieve better predictive performance.

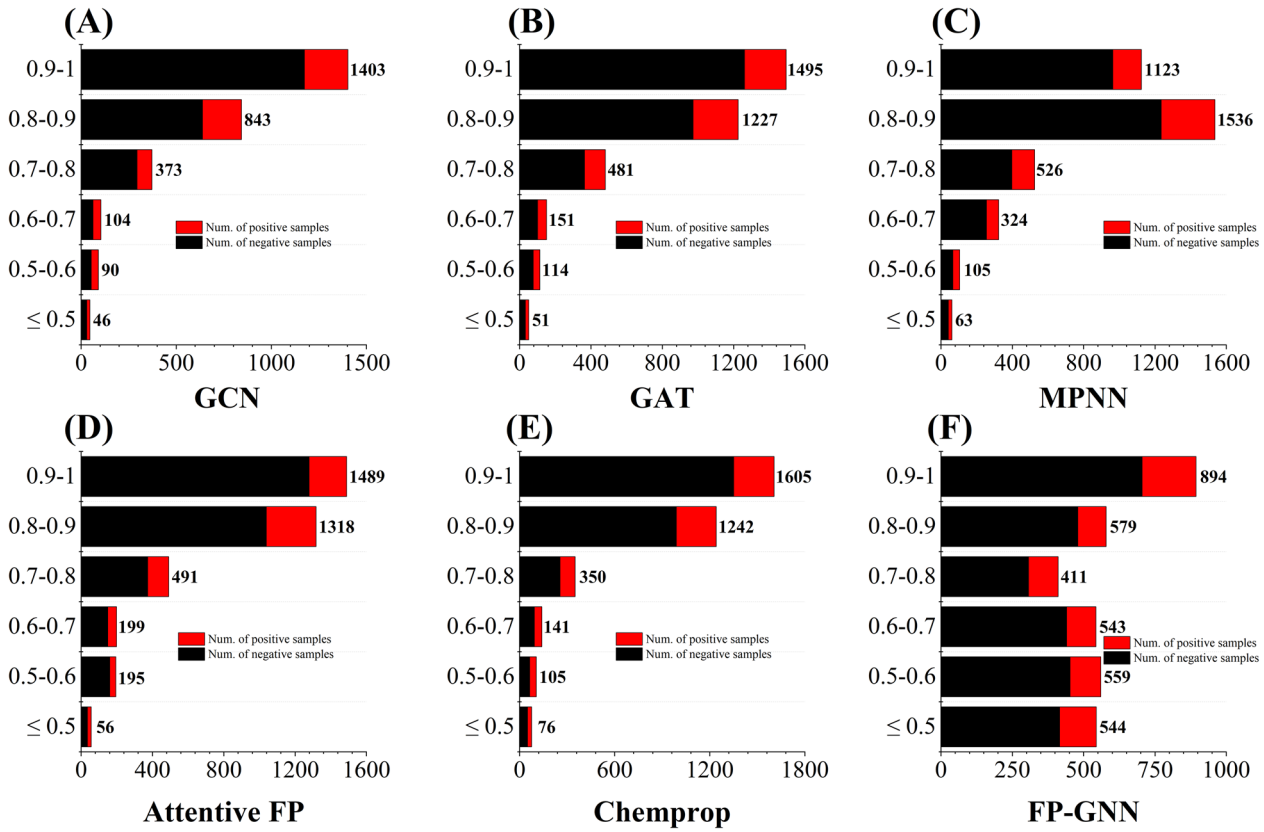

Fig. 7 Relationships between the interval distribution of AUC values in the test sets and the corresponding interval of different compound quantities in the training sets of GCN (**A**), GAT (**B**), MPNN (**C**), Attentive FP (**D**), Chemprop (**E**), and FP-GNN (**F**) models

Comparison performance results of fingerprint-, descriptor-, and graph-based ML and DL models

Boxplots analysis for AUC values of descriptor-, fingerprint-, and graph-based models on the test sets are shown in Fig. 8. If considering the commonly used AUC value as the final evaluation metric, the RF::RDKitDes models (Fig. 8F) performed best, followed by RF::FP2, RF::AtomPairs, and RF::Morgan models. It is clear that RF method usually achieved the best performance (Fig. 8) for the kinase profiling prediction task when molecular descriptors and fingerprints are used as input features. When F1 score, BA and MCC values were used as the final assessment metric (Additional file 1: Figs. S2-S4), RF also showed comparable performance. In addition, the average predictive performance of graph-based DL algorithms (Fig. 8G and Figs. S2G-S4G) are inferior to fingerprints- and descriptor-based ML models. The optimal in silico predictive models for each kinase in terms of AUC metric are shown in Additional file 2: Table S6.

For better comparison of the predictive performance of deep learning to a variety of other prediction methods, based on KinaseNet dataset, we added multi-task GCN, GAT, DNN, FP-GNN, Chemprop and Attentive FP models. A total of six deep learning methods were adopted to construct the corresponding multi-task deep learning models, and hyperparameters optimazation were performed to stretch the ability of algorithms. As shown in Table 5, compared with single model, multi-task learning can promote the comprehensive prediction ability of the model, and improve the prediction ability of models on the multi-task data set. In addition, the multi-task FP-GNN model achieves the highest average AUC of 0.807, which is higher than the best descriptor models (0.798) and fingerprint models (0.786). Besides, the multi-task FP-GNN model's performance is close to but slightly worse than RF::AtomPairs + FP2 + RDKitDes fusion model (0.825). These results show that the effects of descriptor-based and graph-based models vary from data set to data set. Although current research focuses on graph-based multitask modeling strategies, and many graph-based deep learning and multi-task models claim to have the most advanced performance in predictive tasks, there is still much debate about the performance of algorithms based on molecular fingerprints and descriptors versus those based on molecular pictures and structures.

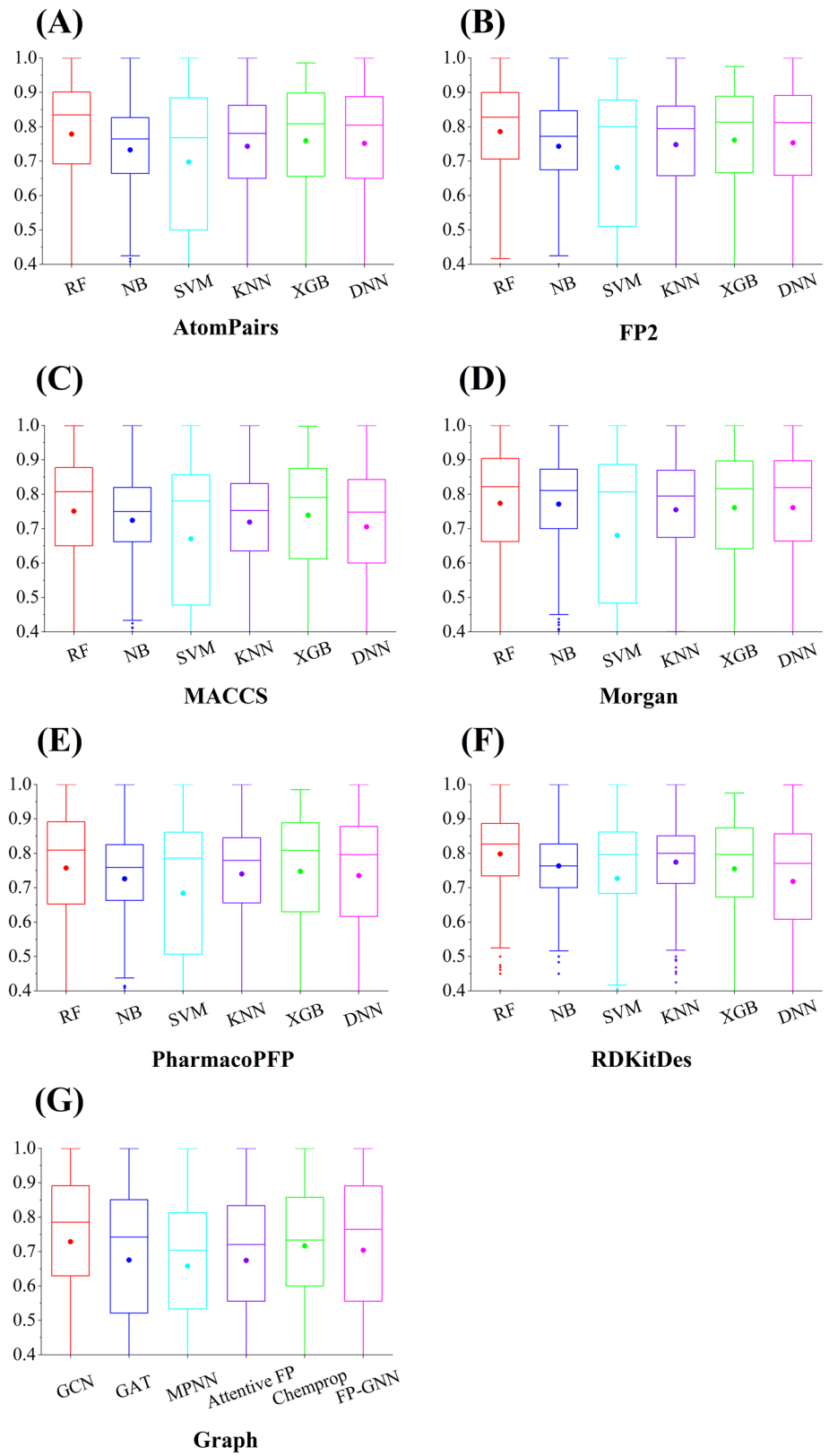

Fig. 8 Comparison of average AUC values of **A** AtomPairs-, **B** FP2-, **C** MACCS-, **D** Morgan-, **E** PharmacoPFP-, and **F** RDKitDes-, **G** Graph-based models using five ML and one DNN DL methods. The average AUC values of the test sets for various ML and DL algorithms are displayed as boxplot. Middle spheres represent the median, and boxes represent the interquartile range (IQR) from the median

Table 5 Performance of AUC values based on multi-task models

Single models	AUROC	Multi-task models	AUROC
GCN	0.729	Multi-GCN	0.785
GAT	0.675	Multi-GAT	0.713
FP-GNN	0.704	FP-GNN	0.807
Chemprop	0.717	Chemprop	0.798
AttentiveFP	0.674	AttentiveFP	0.667
MPNN	0.658	Multi-DNN::Morgan	0.556

Exploring whether combining descriptors and fingerprints could improve the performance of models

To investigate whether the combined features of fingerprints and descriptors could improve the performance of the kinase profiling prediction task, the combined features were used to establish 10,620 models using six ML algorithms. As shown in Table 6, the combined-features-models based on the RF, XGBoost and DNN algorithms slightly outperformed their corresponding descriptor- and fingerprint-based models in terms of AUC metric. For example, the best combined-features-model (RF::Morgan::RDkitDes, AUC = 0.815, Table 6) is superior to RF::RDKitDes and RF::Morgan. Similar trends occurred in the comparative performance of the combined-features-models and individual descriptor- and fingerprint-based models in terms of F1 score (Additional file 2: Table S7). However, the predictive performance of the combined-features-models constructed using KNN, NB, and SVM methods did not outperform the corresponding descriptor-based models (Table 6, because the average AUC values of these combined models were slightly larger than that of the fingerprint-based models, but smaller than that of the descriptor-based models. A possible reason is that more input of feature information is conducive to building accurate prediction models for the ensemble learning RF and XGB algorithms and DNN method.

Exploring whether model fusion could improve performance on the kinase profiling prediction task

We further explore whether fusion models can improve classification accuracy of a single model in the kinase profiling prediction task. Given that the RF, KNN and NB algorithms outperformed other ML and DL methods on the kinase profiling prediction task (Additional file 2: Table S8), both voting and stacking methods were therefore used construct fusion model based on the three ML algorithms. As shown in Fig. 9, both voting- and stacking-based fusion models were slightly better than the corresponding single-based RF and KNN models, albeit with some differences in terms of NB models. For example, the voting fusion models based on RF achieved the best overall performance with the highest average values of AUC (0.825 ± 0.124).

Collectively, RF::AtomPairs + FP2 + RDKitDes voting fusion models achieved the overall best performance in the kinome-wide profiling prediction task in terms of AUC metric. As shown in Fig. 10A, 301 HQ models were obtained in the voting fusion models and distributed over the entire kinome tree covering all kinase families.

KIPP online webserver construction and application

Although several kinases profiling prediction models have been reported (Additional file 2: Table S9), easy-to-use software and/or online webserver are not available. To this end, an online platform called KIPP (https://kipp.idruglab.cn/) was developed based on the overall optimal RF::AtomPairs + FP2 + RDKitDes models (default). A collection of the best models based on each kinase and the multi-task FP-GNN model are also provided. KIPP includes five main modules: compound basic information display, kinase profiling prediction and display, kinase tree construction and display, selectivity index calculation and display, and similarity search results display. Overall selectivity and selectivity towards a kinase subfamily will be generated based on the predicted kinase profile. The overall selectivity is represented by the two quantitative evaluation methods, standard score [84] and Gini coefficient [85]. Odds ratio (OR) is adopted to calculate sub-family selectivity to represent the strength of the association between an inhibitor and a sub-family [86].

Taking CHMFL-BMX-078 (a highly potent and selective Type II irreversible BMX kinase inhibitor) [87] as an example, users can easily upload the SMILES or draw the structure online of CHMFL-BMX-078 (Fig. 10B) to quickly predict the inhibitory activity of this compound against 354 kinase across the kinome. Once the calculation task is completed, users can click on different modules to query the calculation results, including basic compound information (Fig. 11A), kinase profiling prediction results in heat map (Fig. 11B) and list

Table 6 Performance comparison results of AUC values between the combined-features-based models and individual descriptor- and fingerprint-based models

Method	Combined features	AUC	Molecular feature	AUC	Difference
DNN[a]	AtomPairs::RDKitDes	0.749	AtomPairs	0.752	− 0.003
	FP2::RDKitDes	0.762	FP2	0.753	0.009
	MACCS::RDKitDes	0.741	MACCS	0.705	0.036
	Morgan::RDKitDes	0.774	Morgan	0.761	0.013
	PharamacoPFP::RDKitDes	0.748	PharamacoPFP	0.735	0.013
			RDKitDes	0.718	
KNN[b]	AtomPairs::RDKitDes	0.745	AtomPairs	0.743	0.002
	FP2::RDKitDes	0.754	FP2	0.748	0.006
	MACCS::RDKitDes	0.742	MACCS	0.719	0.023
	Morgan::RDKitDes	0.767	Morgan	0.755	0.012
	PharmacoPFP::RDKitDes	0.749	PharmacoPFP	0.740	0.009
			RDKitDes	0.774	
NB[c]	AtomPairs::RDKitDes	0.738	AtomPairs	0.733	0.005
	FP2::RDKitDes	0.747	FP2	0.743	0.004
	MACCS::RDKitDes	0.750	MACCS	0.724	0.026
	Morgan::RDKitDes	0.781	Morgan	0.772	0.009
	PharmacoPFP::RDKitDes	0.737	PharmacoPFP	0.726	0.011
			RDKitDes	0.763	
RF[d]	AtomPairs::RDKitDes	0.792	AtomPairs	0.779	0.013
	FP2::RDKitDes	0.803	FP2	0.786	0.017
	MACCS::RDKitDes	0.799	MACCS	0.751	0.048
	Morgan::RDKitDes	0.815	Morgan	0.774	0.041
	PharmacoPFP::RDKitDes	0.801	PharmacoPFP	0.757	0.044
			RDKitDes	0.798	
SVM[e]	AtomPairs::RDKitDes	0.699	AtomPairs	0.698	0.001
	FP2::RDKitDes	0.686	FP2	0.682	0.004
	MACCS::RDKitDes	0.681	MACCS	0.670	0.011
	Morgan::RDKitDes	0.685	Morgan	0.680	0.005
	PharmacoPFP::RDKitDes	0.687	PharmacoPFP	0.684	0.003
			RDKitDes	0.727	
XGBoost[f]	AtomPairs::RDKitDes	0.763	AtomPairs	0.759	0.004
	FP2::RDKitDes	0.768	FP2	0.761	0.007
	MACCS::RDKitDes	0.758	MACCS	0.739	0.019
	Morgan::RDKitDes	0.768	Morgan	0.761	0.007
	PharmacoPFP::RDKitDes	0.763	PharmacoPFP	0.748	0.015
			RDKitDes	0.755	

[a] DNN: Deep neural networks

[b] KNN: K-Nearest Neighbor

[c] NB: Naïve Bayesian

[d] RF: Random forest

[e] SVM: Support vector machine

[f] XGBoost: Extreme gradient boosting

(Fig. 11C), kinase tree diagram (Fig. 11D), selectivity index results (Fig. 11E) and similarity search results for the CHMFL-BMX-078 (Fig. 11F). The predicted kinase profiling results of CHMFL-BMX-078 by KIPP were overall consistent with the experimental kinases inhibition results (Additional file 2: Table S10), with an AUC

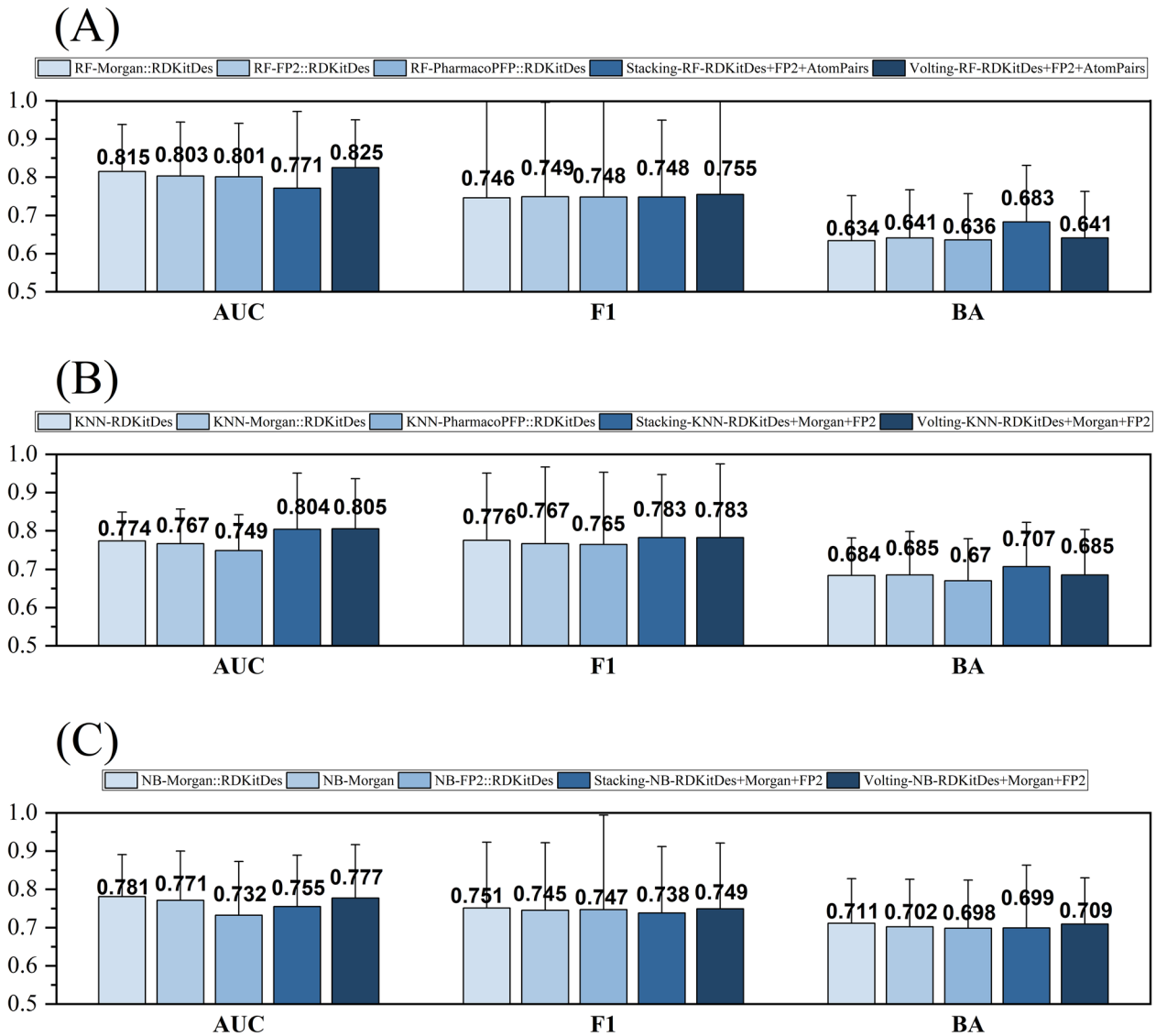

Fig. 9 Comparison of the prediction results between fusion models and single models. The fusion models are constructed based on RF (**A**), KNN (**B**), and NB (**C**) models using voting and stacking methods

value of 0.763, indicating the accuracy and usability of the KIPP platform. Importantly, native versions of Python software are also provided for various kinases, allowing users to perform large-scale VS.

Conclusions

In this paper, we provided a comprehensive assessment of the performance of five ML (NB, RF, XGBoost, KNN, and SVM) and seven DL (DNN, GCN, GAT, MPNN, D-MPNN, Attentive FP, and FP-GNN) methods in kinase profiling prediction task. To obtain a more objective performance evaluation, we constructed a comprehensive KinaseNet dataset covering 354 kinases cross the entire kinome to benchmark all tools. Three types of commonly used molecular features, including a set of molecular descriptors, a collection of five molecular fingerprints (Morgan, MACCS keys, AtomParis, FP2, and PharmacoPFP), and molecular graphs, were used as input features to build predictive models using these compared methods. We found that RF outperform the other

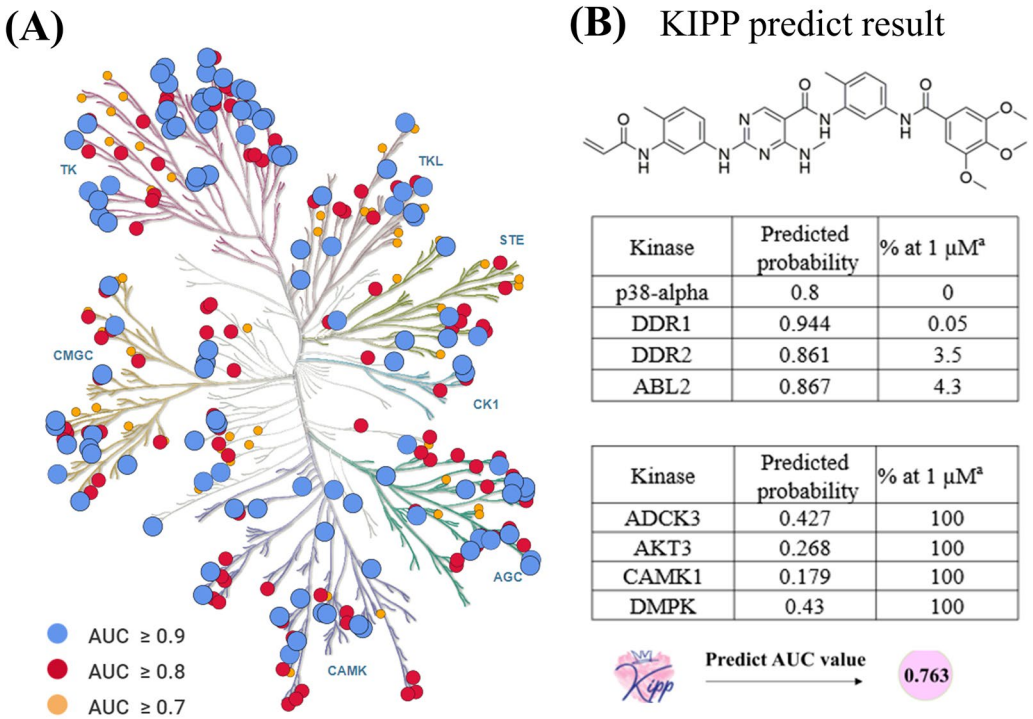

Fig. 10 **A** Kinome map analysis of the RF::AtomPairs + FP2 + RDKitDes models. Kinases are colored based on their AUC values. The kinase tree was generated using Kinmap tool (http://kinhub.org/kinmap) [83]. **B** Chemical structure of CHMFL-BMX-078 and its predicted result. AUC value (0.763) was generated based on the predicted kinase profile of CHMFL-BMX-078 using KIPP and its experimentally tested kinase profile. BMX: bone marrow kinase in the X chromosome

methods for kinase profiling prediction. This finding generalizes across different types of molecular descriptors and fingerprints. Meanwhile, the RDKitDes-based models generally outperform fingerprint-based models. Specifically, the RF::RDKitDes models performed best, followed by RF::FP2, RF::AtomPairs, and RF::Morgan models. Although single-task graph-based DL methods do not achieve the best overall predictive performance on the KinaseNet dataset, the predictive performance of multi-task DL models such as multitask FP-GNN and Chemprop models can still achieve comparable or even better predictive performance than conventional descriptor- and fingerprint-based models, due to the existence of certain data linkages between the various kinase data. In addition, these performance of DL methods improves as the training dataset increases. Accordingly, we envision that with the increasing amounts and quality of data from industry and academia, further performance improvements could be gained by DL methods. Combining descriptors and fingerprints could improve the performance of models, especially for the fingerprint-based models. In addition, fusion models based on the voting and stacking methods further improve performance on the kinase profiling prediction task. Finally, an easy-to-use online platform KIPP and its local version software were constructed based on the optimal models for various kinase inhibitor identification related tasks, including kinase profiling prediction, virtual screening, drug repositioning, and target fishing. It is expected that this study can provide valuable guidance for researchers who are interested in developing innovative and even more powerful kinase profiling prediction models, as well as for medicinal chemists and pharmacologists in designing and discovering new kinase inhibitors.

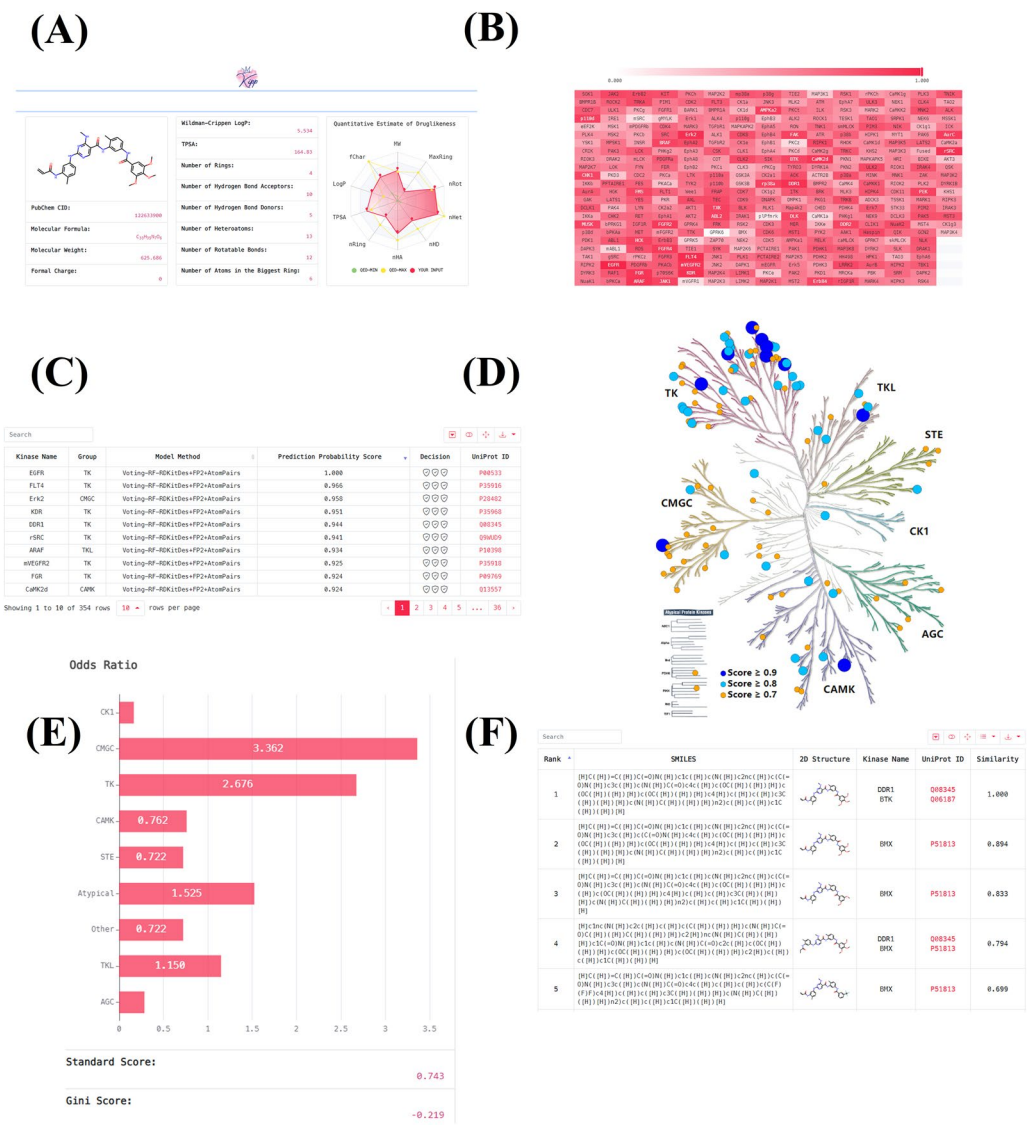

Fig. 11 Website schematic diagram of KIPP for CHMFL-BMX-078 in the kinase profiling prediction task. **A** represents the basic information for submitted CHMFL-BMX-078. **B** and **C** represent kinase profiling prediction results of CHMFL-BMX-078 in heat map and list, respectively. **D** represents kinase tree diagram of CHMFL-BMX-078. **E** represents the selectivity index results. **F** represents the similarity search results

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-023-00799-5.

Additional file 1: Fig S1. Detailed comparison performance of descriptor- and fingerprint-based models using various ML algorithms. (A), (B), (C), (D), (E), and (F) represent the comparison results for the RF, NB, SVM, KNN, XGB, and DNN methods, respectively. **Fig S2.** Comparison of average F1 scores of (A) AtomPairs-, (B) FP2-, (C) MACCS-, (D) Morgen-, (E) PharmacoPFP-, (F) RDKitDes-, and (G) Graph-based models. The assay-F1 scores for various ML algorithms are displayed as boxplot. Middle spheres represent the median, and boxes represents the interquartile range (IQR) from the median. **Fig S3.** Comparison of average BA values of (A) AtomPairs-, (B) FP2-, (C) MACCS-, (D) Morgen-, (E) PharmacoPFP-, (F) RDKitDes-, and (G) Graph-based models. The assay-BA values for various ML algorithms are displayed as boxplot. Middle spheres represent the median, and boxes represent the interquartile range (IQR) from the median. **Fig S4.** Comparison of average MCC values of (A) AtomPairs-, (B) FP2-, (C) MACCS-, (D) Morgen-, (E) PharmacoPFP-, and (F) RDKitDes-, (G) Graph-based models. The assay-MCC values for various ML algorithms are displayed as boxplot. Middle spheres represent the median, and boxes represent the interquartile range (IQR) from the median.

Additional file 2: Table S1. Details on benchmark dataset for kinase profiling prediction task used in this study. **Table S2.** Structural diversity and chemical space analysis of the compounds in each kinase. **Table S3.** Detailed performance results of different ML methods. **Table S4.** Detailed individual kinases where the GCN models outperform the RF::RDKitDes models. **Table S5.** Detailed individual kinases where the FP-GNN models outperform the RF::RDKitDes models. **Table S6.** The optimal in silico predictive models for each kinase in terms of AUC metric. **Table S7.** Comparison performance of models based on combined features and single feature in terms of F1 score.

Table S8. Ranking of all single models by AUC values. **Table S9.** Comparison of our models with the reported in silico prediction models for kinase profiling prediction task. **Table S10.** The predicted activity probability and experimental % activity of CHMFL-BMX-078.

Acknowledgements
We acknowledge the allocation time from the SCUTGrid at South China University of Technology.

Author contributions
LW conceived and designed the project. JW and YC contributed to the literature search, data collection, and algorithm architecture realization. JW and DZ was response for analyzing the modelling results and implementation models to web-server. JH and ML were in charge of web-based software construction on front-end and back-end respectively. LW provided support and critically revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
This work was supported in part by the National Natural Science Foundation of China (81973241) and the Natural Science Foundation of Guangdong Province (2020A1515010548).

Data availability
KIPP online platform is freely accessible at https://kipp.idruglab.cn/. Datasets and python version executable software of KIPP are freely available on Github: https://github.com/idrugLab/KinasePredictPro.

Declarations

Competing interests
The authors declare no competing interests.

Received: 30 May 2023 Accepted: 22 December 2023
Published online: 30 January 2024

References

1. Manning G, Whyte DB, Martinez R et al (2002) The protein kinase complement of the human genome. Science 298:1912–1934. https://doi.org/10.1126/science.1075762
2. Huang M, Shen A, Ding J, Geng M (2014) Molecularly targeted cancer therapy: some lessons from the past decade. Trends Pharmacol Sci 35:41–50. https://doi.org/10.1016/j.tips.2013.11.004
3. Ma WW, Adjei AA (2009) Novel agents on the horizon for cancer therapy. CA Cancer J Clin 59:111–137. https://doi.org/10.3322/caac.20003
4. Sun C, Bernards R (2014) Feedback and redundancy in receptor tyrosine kinase signaling: relevance to cancer therapies. Trends Biochem Sci 39:465–474. https://doi.org/10.1016/j.tibs.2014.08.010
5. Clark JD, Flanagan ME, Telliez J-B (2014) Discovery and development of janus kinase (JAK) inhibitors for inflammatory diseases: miniperspective. J Med Chem 57:5023–5038. https://doi.org/10.1021/jm401490p
6. Barnes PJ (2013) New anti-inflammatory targets for chronic obstructive pulmonary disease. Nat Rev Drug Discov 12:543–559. https://doi.org/10.1038/nrd4025
7. Muth F, Günther M, Bauer SM et al (2015) Tetra-substituted pyridinylimidazoles as dual inhibitors of p38α mitogen-activated protein kinase and c-Jun N-terminal kinase 3 for potential treatment of neurodegenerative diseases. J Med Chem 58:443–456. https://doi.org/10.1021/jm501557a
8. Kikuchi R, Nakamura K, MacLauchlan S et al (2014) An antiangiogenic isoform of VEGF-A contributes to impaired vascularization in peripheral artery disease. Nat Med 20:1464–1471. https://doi.org/10.1038/nm.3703
9. Banks AS, McAllister FE, Camporez JPG et al (2015) An ERK/Cdk5 axis controls the diabetogenic actions of PPARγ. Nature 517:391–395. https://doi.org/10.1038/nature13887
10. Nygaard HB, van Dyck CH, Strittmatter SM (2014) Fyn kinase inhibition as a novel therapy for Alzheimer's disease. Alzheimers Res Ther 6:8. https://doi.org/10.1186/alzrt238
11. Attwood MM, Fabbro D, Sokolov AV et al (2021) Author correction: trends in kinase drug discovery: targets, indications and inhibitor design. Nat Rev Drug Discov. https://doi.org/10.1038/s41573-021-00303-4
12. Goldstein DM, Gray NS, Zarrinkar PP (2008) High-throughput kinase profiling as a platform for drug discovery. Nat Rev Drug Discov 7:391–397. https://doi.org/10.1038/nrd2541
13. Li D-D, Meng X-F, Wang Q et al (2018) Consensus scoring model for the molecular docking study of mTOR kinase inhibitor. J Mol Graph Model 79:81–87. https://doi.org/10.1016/j.jmgm.2017.11.003
14. Burggraaff L, Lenselink EB, Jespers W et al (2020) Successive statistical and structure-based modeling to identify chemically novel kinase inhibitors. J Chem Inf Model 60:4283–4295. https://doi.org/10.1021/acs.jcim.9b01204
15. Kothiwale S, Borza C, Pozzi A, Meiler J (2017) Quantitative structure-activity relationship modeling of kinase selectivity profiles. Molecules 22:1576. https://doi.org/10.3390/molecules22091576
16. Kong Y, Yan A (2017) QSAR models for predicting the bioactivity of Polo-like kinase 1 inhibitors. Chemom Intell Lab Syst 167:214–225. https://doi.org/10.1016/j.chemolab.2017.06.011
17. Sciabola S, Stanton RV, Wittkopp S et al (2008) Predicting kinase selectivity profiles using free-Wilson QSAR analysis. J Chem Inf Model 48:1851–1867. https://doi.org/10.1021/ci800138n
18. Sheridan RP, Nam K, Maiorov VN et al (2009) QSAR models for predicting the similarity in binding profiles for pairs of protein kinases and the variation of models between experimental data sets. J Chem Inf Model 49:1974–1985. https://doi.org/10.1021/ci900176y
19. Hillisch A, Heinrich N, Wild H (2015) Computational chemistry in the pharmaceutical industry: from childhood to adolescence. ChemMedChem 10:1958–1962. https://doi.org/10.1002/cmdc.201500346
20. Keiser MJ, Roth BL, Armbruster BN et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206. https://doi.org/10.1038/nbt1284
21. Keiser MJ, Setola V, Irwin JJ et al (2009) Predicting new molecular targets for known drugs. Nature 462:175–181. https://doi.org/10.1038/nature08506
22. Martin E, Mukherjee P, Sullivan D, Jansen J (2011) Profile-QSAR: a novel meta-qsar method that combines activities across the kinase family to accurately predict affinity, selectivity, and cellular activity. J Chem Inf Model 51:1942–1956. https://doi.org/10.1021/ci1005004
23. Xia X, Maliski EG, Gallant P, Rogers D (2004) Classification of kinase inhibitors using a bayesian model. J Med Chem 47:4463–4470. https://doi.org/10.1021/jm0303195
24. Schürer SC, Muskal SM (2013) Kinome-wide activity modeling from diverse public high-quality data sets. J Chem Inf Model 53:27–38. https://doi.org/10.1021/ci300403k
25. Lapins M, Wikberg JE (2010) Kinome-wide interaction modelling using alignment-based and alignment-independent approaches for kinase description and linear and non-linear data analysis techniques. BMC Bioinformatics 11:339. https://doi.org/10.1186/1471-2105-11-339
26. Niijima S, Shiraishi A, Okuno Y (2012) Dissecting kinase profiling data to predict activity and understand cross-reactivity of kinase inhibitors. J Chem Inf Model 52:901–912. https://doi.org/10.1021/ci200607f
27. Chen B, Sheridan RP, Hornak V, Voigt JH (2012) Comparison of random forest and pipeline pilot naïve bayes in prospective QSAR predictions. J Chem Inf Model 52:792–803. https://doi.org/10.1021/ci200615h
28. Cao D-S, Zhou G-H, Liu S et al (2013) Large-scale prediction of human kinase–inhibitor interactions using protein sequences and molecular topological structures. Anal Chim Acta 792:10–18. https://doi.org/10.1016/j.aca.2013.07.003
29. Bora A, Avram S, Ciucanu I et al (2016) Predictive models for fast and effective profiling of kinase inhibitors. J Chem Inf Model 56:895–905. https://doi.org/10.1021/acs.jcim.5b00646
30. Merget B, Turk S, Eid S et al (2017) Profiling prediction of kinase inhibitors: toward the virtual assay. J Med Chem 60:474–485. https://doi.org/10.1021/acs.jmedchem.6b01611
31. Yabuuchi H, Niijima S, Takematsu H et al (2011) Analysis of multiple compound–protein interactions reveals novel bioactive molecules. Mol Syst Biol 7:472. https://doi.org/10.1038/msb.2011.5

32. Unterthiner T, Mayr A, Klambauer G, et al. Deep Learning as an Opportunity in Virtual Screening. In: Workshop on Deep Learning and Representation Learning (NIPS2014). 2014.
33. Li X, Li Z, Wu X et al (2020) Deep learning enhancing kinome-wide polypharmacology profiling: model construction and experiment validation. J Med Chem 63:8723–8737. https://doi.org/10.1021/acs.jmedchem.9b00855
34. Avram S, Bora A, Halip L, Curpăn R (2018) Modeling kinase inhibition using highly confident data sets. J Chem Inf Model 58:957–967. https://doi.org/10.1021/acs.jcim.7b00729
35. Li B, Lin M, Chen T, Wang L (2023) FG-BERT: a generalized and self-supervised functional group-based molecular representation learning framework for properties prediction. Brief Bioinform 24:bbad398. https://doi.org/10.1093/bib/bbad398
36. Wu Z, Jiang D, Hsieh C-Y et al (2021) Hyperbolic relational graph convolution networks plus: a simple but highly efficient QSAR-modeling method. Brief Bioinform 22:bbab112. https://doi.org/10.1093/bib/bbab112
37. Ye Q, Chai X, Jiang D et al (2021) Identification of active molecules against Mycobacterium tuberculosis through machine learning. Brief Bioinform 22:bbab068. https://doi.org/10.1093/bib/bbab068
38. Luukkonen S, Meijer E, Tricarico GA et al (2023) Large-scale modeling of sparse protein kinase activity data. J Chem Inf Model 63:3688–3696. https://doi.org/10.1021/acs.jcim.3c00132
39. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13:21–27. https://doi.org/10.1109/TIT.1967.1053964
40. Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley, Hoboken. https://doi.org/10.1007/978-1-4471-0285-4
41. Zernov VV, Balakin KV, Ivaschenko AA et al (2003) Drug discovery using support vector machines. the case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. J Chem Inf Comput Sci 43:2048–2056. https://doi.org/10.1021/ci0340916
42. Svetnik V, Liaw A, Tong C et al (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci 43:1947–1958. https://doi.org/10.1021/ci034160g
43. Chen T, Guestrin C. Xgboost: A scalable tree boosting system//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785–794. https://doi.org/10.1145/2939672.2939785
44. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133. https://doi.org/10.1007/BF02478259
45. Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv. 2017; 160902907
46. Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks. arXiv. 2018; 171010903
47. Gilmer J, Schoenholz SS, Riley PF, et al. Neural message passing for Quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. JMLR.org, Sydney, NSW, Australia, pp 1263–1272. 2017.
48. Xiong Z, Wang D, Liu X et al (2020) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem 63:8749–8760. https://doi.org/10.1021/acs.jmedchem.9b00959
49. Yang K, Swanson K, Jin W et al (2019) Analyzing learned molecular representations for property prediction. J Chem Inf Model 59:3370–3388. https://doi.org/10.1021/acs.jcim.9b00237
50. Cai H, Zhang H, Zhao D et al (2022) FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. Brief Bioinform 23(6):bbac408
51. Mendez D, Gaulton A, Bento AP et al (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 47:D930–D940. https://doi.org/10.1093/nar/gky1075
52. Kim S, Chen J, Cheng T et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49:D1388–D1395. https://doi.org/10.1093/nar/gkaa971
53. Liu T, Lin Y, Wen X et al (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res 35:D198–D201. https://doi.org/10.1093/nar/gkl999
54. Sterling T, Irwin JJ (2015) ZINC 15—ligand discovery for everyone. J Chem Inf Model 55:2324–2337. https://doi.org/10.1021/acs.jcim.5b00559
55. Laufkötter O, Laufer S, Bajorath J (2020) Kinase inhibitor data set for systematic analysis of representative kinases across the human kinome. Data Brief 32:106189. https://doi.org/10.1016/j.dib.2020.106189
56. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50:742–754. https://doi.org/10.1021/ci100050t
57. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci 42:1273–1280. https://doi.org/10.1021/ci010132r
58. Carhart RE, Smith DH, Venkataraghavan R (1985) Atom pairs as molecular features in structure-activity studies: definition and applications. J Chem Inf Comput Sci 25:64–73. https://doi.org/10.1021/ci00046a002
59. O'Boyle NM, Banck M, James CA et al (2011) Open babel: an open chemical toolbox. J Cheminform 3:33. https://doi.org/10.1186/1758-2946-3-33
60. Gobbi A, Poppinger D (1998) Genetic optimization of combinatorial libraries. Biotechnol Bioeng 61:47–54. https://doi.org/10.1002/(SICI)1097-0290(199824)61:1%3c47::AID-BIT9%3e3.0.CO;2-Z
61. Kearnes S, McCloskey K, Berndl M et al (2016) Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des 30:595–608. https://doi.org/10.1007/s10822-016-9938-8
62. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, et al. Convolutional Networks on Graphs for Learning Molecular Fingerprints. arXiv. 2015; 150909292
63. Wang L, Le X, Li L et al (2014) Discovering new agents active against methicillin-resistant staphylococcus aureus with ligand-based approaches. J Chem Inf Model 54:3186–3197. https://doi.org/10.1021/ci500253q
64. Wang L, Chen L, Yu M et al (2016) Discovering new mTOR inhibitors for cancer treatment through virtual screening methods and in vitro assays. Sci Rep 6:18987. https://doi.org/10.1038/srep18987
65. Luo Y, Zeng R, Guo Q et al (2019) Identifying a novel anticancer agent with microtubule-stabilizing effects through computational cell-based bioactivity prediction models and bioassays. Org Biomol Chem 17:1519–1530. https://doi.org/10.1039/c8ob02193g
66. Guo Q, Zhang H, Deng Y et al (2020) Ligand- and structural-based discovery of potential small molecules that target the colchicine site of tubulin for cancer treatment. Eur J Med Chem 196:112328. https://doi.org/10.1016/j.ejmech.2020.112328
67. Joachims T. Text categorization with support vector machines : learning with many relevant features. Proceedings of the ECML-98. 1998.
68. Li S, Ding Y, Chen M et al (2021) HDAC3i-finder: a machine learning-based computational tool to screen for HDAC3 inhibitors. Mol Inform 40:2000105. https://doi.org/10.1002/minf.202000105
69. Jiang D, Wu Z, Hsieh C-Y et al (2021) Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. J Cheminformatics 13:12. https://doi.org/10.1186/s13321-020-00479-8
70. Gawehn E, Hiss JA, Schneider G (2016) Deep learning in drug discovery. Mol Inform 35:3–14. https://doi.org/10.1002/minf.201501008
71. Ma J, Sheridan RP, Liaw A et al (2015) Deep neural nets as a method for quantitative structure-activity relationships. J Chem Inf Model 55:263–274. https://doi.org/10.1021/ci500747n
72. Zhu L, Wan B, Li C et al (2021) Dyadic relational graph convolutional networks for skeleton-based human interaction recognition. Pattern Recognit 115:107920. https://doi.org/10.1016/j.patcog.2021.107920
73. Flam-Shepherd D, Wu T, Friederich P, Aspuru-Guzik A. Neural message passing on high order paths. arXiv. 2020; 200210413
74. Withnall M, Lindelöf E, Engkvist O, Chen H (2020) Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction. J Cheminform 12:1. https://doi.org/10.1186/s13321-019-0407-y
75. Tang B, Kramer ST, Fang M et al (2020) A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility. J Cheminform 12:15. https://doi.org/10.1186/s13321-020-0414-z
76. Stokes JM, Yang K, Swanson K et al (2020) A deep learning approach to antibiotic discovery. Cell 180:688-702.e13. https://doi.org/10.1016/j.cell.2020.01.021
77. Swami A, Jain R (2013) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

78. Sorgenfrei FA, Fulle S, Merget B (2018) Kinome-wide profiling prediction of small molecules. ChemMedChem 13:495–499. https://doi.org/10.1002/cmdc.201700180
79. Abdelbaky I, Tayara H, Chong KT (2021) Prediction of kinase inhibitors binding modes with machine learning and reduced descriptor sets. Sci Rep 11:706. https://doi.org/10.1038/s41598-020-80758-4
80. Sánchez-Cruz N, Medina-Franco JL (2021) Epigenetic target fishing with accurate machine learning models. J Med Chem 64:8208–8220. https://doi.org/10.1021/acs.jmedchem.1c00020
81. Kc GB, Bocci G, Verma S et al (2021) A machine learning platform to estimate anti-SARS-CoV-2 activities. Nat Mach Intell 3:527–535. https://doi.org/10.1038/s42256-021-00335-w
82. Wu Z, Ramsundar B, Feinberg EN et al (2018) MoleculeNet: a benchmark for molecular machine learning. Chem Sci 9:513–530. https://doi.org/10.1039/C7SC02664A
83. Eid S, Turk S, Volkamer A et al (2017) KinMap: a web-based tool for interactive navigation through human kinome data. BMC Bioinformatics 18:1–6
84. Karaman MW, Herrgard S, Treiber DK et al (2008) A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol 26:127–132. https://doi.org/10.1038/nbt1358
85. Graczyk PP (2007) Gini coefficient: a new way to express selectivity of kinase inhibitors against a family of kinases. J Med Chem 50:5773–5779. https://doi.org/10.1021/jm070562u
86. Bland JM (2000) Statistics notes: the odds ratio. BMJ 320:1468–1468. https://doi.org/10.1136/bmj.320.7247.1468
87. Liang X, Lv F, Wang B et al (2017) Discovery of 2-((3-Acrylamido-4-methylphenyl)amino)-N-(2-methyl-5-(3,4,5-trimethoxybenzamido)phenyl)-4-(methylamino)pyrimidine-5-carboxamide (CHMFL-BMX-078) as a highly potent and selective type II irreversible bone marrow kinase in the X chromosome (BMX) kinase inhibitor. J Med Chem 60:1793–1816. https://doi.org/10.1021/acs.jmedchem.6b01413

Publisher's Note