

RESEARCH ARTICLE

Open Access



A multitask GNN-based interpretable model for discovery of selective JAK inhibitors

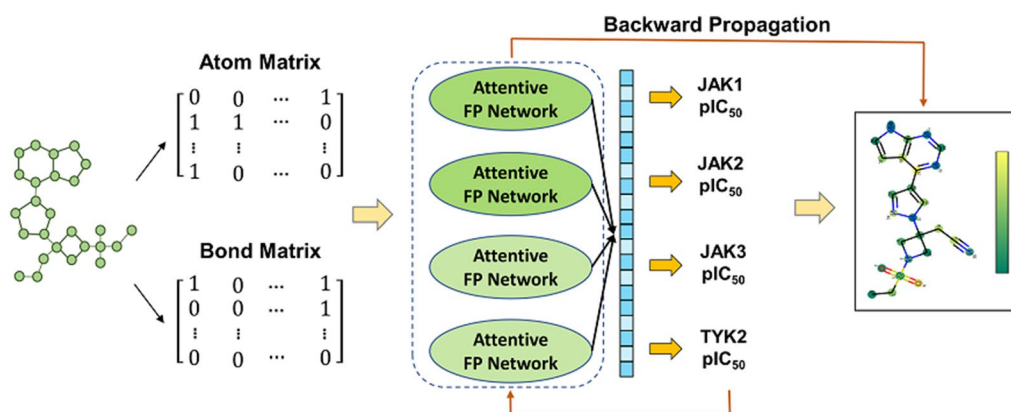
Yimeng Wang, Yaxin Gu, Chaofeng Lou, Yuning Gong, Zengrui Wu, Weihua Li, Yun Tang* and Guixia Liu*

Abstract

The Janus kinase (JAK) family plays a pivotal role in most cytokine-mediated inflammatory and autoimmune responses via JAK/STAT signaling, and administration of JAK inhibitors is a promising therapeutic strategy for several diseases including COVID-19. However, to screen and design selective JAK inhibitors is a daunting task due to the extremely high homology among four JAK isoforms. In this study, we aimed to simultaneously predict pIC_{50} values of compounds for all JAK subtypes by constructing an interpretable GNN multitask regression model. The final model performance was positive, with R^2 values of 0.96, 0.79 and 0.78 on the training, validation and test sets, respectively. Meanwhile, we calculated and visualized atom weights, followed by the rank sum tests and local mean comparisons to obtain key atoms and substructures that could be fine-tuned to design selective JAK inhibitors. Several successful case studies have demonstrated that our approach is feasible and our model could learn the interactions between proteins and small molecules well, which could provide practitioners with a novel way to discover and design JAK inhibitors with selectivity.

Keywords: Selective JAK inhibitors, GNN, Multitask learning, Model interpretations, Key substructures, Applicability domain

Graphical Abstract



Introduction

Cytokines are small proteins produced and secreted by immune and non-immune cells that are involved in

*Correspondence: ytang234@ecust.edu.cn; gxliu@ecust.edu.cn
Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

intercellular signaling and interactions [1, 2]. They promote and restrict one another, forming an extremely complex cytokine immune regulatory network [3]. When multiple cytokines released, the balance of the network would be broken, leading to the loss of immune homeostasis and causing a few immune-mediated inflammatory diseases [3]. More seriously, it would result in a cytokine storm that often happened in viral infections like severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [4–7]. Cytokine levels could be regulated through the Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway [8]. When cytokine binds to the receptor outside the cell membrane, the cytokine receptor is activated and phosphorylates the JAK and the downstream molecule STAT. The complex then enters into the nucleus and controls the transcription of cellular genes, which in turn impacts the biological function of the cell [9, 10]. Hence, we could design certain JAK inhibitors for the treatment of immune and inflammatory diseases that were mediated by cytokines.

JAK is a class of non-receptor tyrosine kinases, including four subtypes, namely JAK1, JAK2, JAK3, and TYK2, which regulate different cytokines. Currently, nine JAK inhibitors have been approved by the US Food and Drug Administration (FDA), the European Medical Agency or other regulatory agencies (Additional file 1: Table S1). Nonetheless, a large portion of them are non-selective, having more or less toxicities or other undesirable side effects. For example, tofacitinib is the first pan-JAK inhibitor that targets JAK1, JAK2 and JAK3 for the treatment of moderate or severe active ulcerative colitis (UC), but owing to the simultaneous inhibition of JAK2-mediated erythropoietin and GM-CSF signaling, it has some safety concerns relating to significant adverse reactions including anemia, neutropenia, thrombocytopenia, zoster, and pulmonary embolism [11–13]. The similar story also happened in baricitinib, 10% of patients treated with baricitinib showed side effects such as upper respiratory tract infections and high blood cholesterol levels (hypercholesterolemia) [14]. Therefore, it is reasonable to propose that inhibitors targeting one specific JAK isoenzymes without affecting other JAK-dependent signals could spare toxicities and maximize the clinical benefit. However, discovering small molecules that bind selectively to a given protein target has long been a major stumbling block to modern drug discovery [15]. The promise of *in silico* screening is tantalizing, as it would allow compounds to be screened at greatly reduced cost [16].

The common structure of all JAKs consists of four structural domains composed of seven homologous regions [JH1–7] [17]. Most of known small molecular

inhibitors targeting JAKs are active site-directed. They bind to the adenosine triphosphate (ATP) site of the catalytic domain (also referred to the JH1 or “Janus Homology 1” domain) [18]. The crystal structures of the JH1 domains have been resolved for the four JAK isoenzymes. However, due to the high sequential homology and structural similarity of the ATP active site across JAK family, it is too hard to discover highly selective molecules for a specific JAK family member by structure-based virtual screening (VS) methods, as Bajusz’s study did [19]. In recent years, deep learning, a branch of machine learning, has been an effective tool for drug discovery, especially in molecular property prediction and ligand-based VS fields [20]. Compared with traditional machine learning methods, deep learning encompasses several layers of stacked complex neural networks that can represent and learn deeper knowledge [21]. Graph neural networks (GNN) have been gaining popularity among many scholars recently. It is plausible for GNN to represent atoms and bonds with nodes and edges, respectively, which are much more powerful at capturing the latent patterns and require less feature engineering efforts [22]. However, deep learning methods required large datasets and are poorly interpretable. To address these issues, we constructed a multitask regression model based on the attentive fingerprint framework (MTATFP) that allows for the simultaneous prediction of IC_{50} values of compounds for the four JAK isoforms (shown in Fig. 1). In addition, we utilized attention coefficients to assign weights to each atom of the compound and visualize them. Our virtual screening platform could facilitate the discovery and modification of JAK inhibitors.

Methods

Data collection and processing

In this work, the inhibitors of four JAK isoforms and their experimental IC_{50} values were collected from three databases, PubChem [23], ChEMBL [24] and BindingDB [25] (indicated in Fig. 1). Molecules from different sources were converted to SMILES strings and preprocessed by RDKit [26] and MolVS (<https://molvs.readthedocs.io/en/latest/index.html>), including normalization of structures, desalting, neutralization of charge and elimination of duplicate molecules (the canonical SMILES strings were chosen as the unique identification and the repeated molecules with lower IC_{50} values were kept). In order to ensure numerical stability during model building, we transformed IC_{50} (nM) into its negative logarithmic scale pIC_{50} ($-LgIC_{50}$).

For developing multitask regression models, the obtained data set was randomly divided into training set, validation set and test set at a ratio of 8:1:1 by python

scripts. Training set is used to build the model, validation set is used for hyper-parameter optimization and test set is for the model evaluation. Furthermore, to evaluate the generalization ability of our model, we extracted 33 K_d values of JAKs from Davis's works [27] and 152 activity values from Anastassiadis *et al* [28] (both outside the training set) as two external validation sets. In Anastassiadis's datasets, we converted the activity values to IC_{50} values with an equation defined earlier [29].

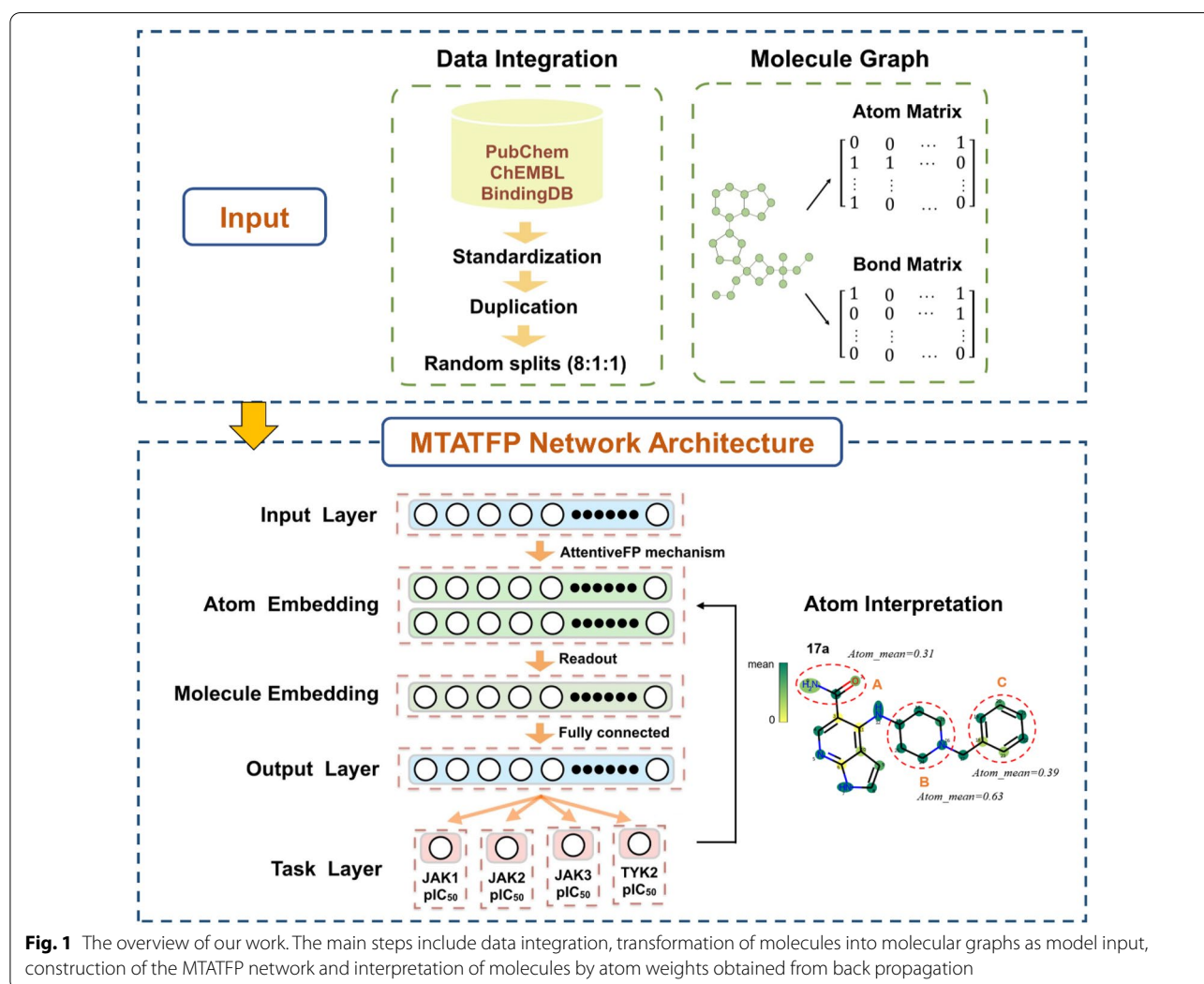
Multitask learning based on attentive FP network

Attentive FP network architecture

As a typical graph convolutional network, graph attention network (GAT) introduced an attention mechanism which has been successfully applied for achieving better neighbor aggregation [30]. The Attentive FP framework was first proposed by Xiong and colleagues [31]. It introduced the graph-attention mechanism to learn both

atomic and molecular properties of a given chemical structure, which is tailored for molecular feature extractions. In the present study, we applied the Attentive FP network architecture on four JAK-based tasks to predict the pIC_{50} values (see Fig. 1). We have constructed two Attentive FP convolutional layers for extracting atomic features and a readout layer for molecular embedding, which ultimately output the predicted values using a fully connected layer. A LeakyReLU [32] function was introduced as a nonlinear activation function after a linear transformation.

Before building the attentive FP network, we translated the JAK-related molecules into molecular graphs using RDKit (version 2020_03_1), the rules were shown in Additional file 1: Table S2.



Multitask learning strategy

In this work, we employed a multitask learning strategy to improve the predictive performance of small datasets. Four tasks shared the same hidden layers and hyper-parameters during training process, and separated to different tasks at the output of the fully connected layer in the network (shown in Fig. 1). It takes advantage of implicit data augmentation by borrowing similarly structured information features [33]. For example, we could allow model eavesdrop to send JAK2-task to learn some features which were difficult for TYK2-task. Furthermore, it helped to focus the model's attention on impactful features since other tasks can provide additional evidence whether the features are relevant or not [34]. Meanwhile, we also constructed four single-task models on each JAK subtype for comparison with the multitask model.

Model training protocol

Concerning the overwhelming complexity and high computational cost of neural networks, we used random searching strategy based on previous experience for hyper-parameter settings, including a few common ones (such as learning rate, weight decay, batch size) and Attentive FP hyper-parameters (such as the number of network layers, graph feat size and dropout). The performance metrics of validation set were used for model selection. To avoid overfitting and saving computational resources, we used early stopping approaches. A maximum epoch was set as 1000. If the performance metric had not improved in 20 epochs on the validation set, the training process was terminated early. Attentive FP was trained by the Deep Graph Library Python (DGL) package (version 0.6.0) [35] with cuda 10.1 and the DGL-LifeSci [36] extension (github.com/aws-labs/dgl-livesci), which ran on the GPU version of the PyTorch [37] framework (version 1.5.0).

Atom interpretation

Lack of interpretability is another issue concerned by machine learning and deep learning, usually called "the black-box mode" [38]. Inversion of the Attentive FP model by extracting the hidden layers or attention weights could provide access to the model's interpretation, which would help chemists gain insights into the skyrocketing volume and complexity of drug discovery data. In the Attentive FP network, since atoms are treated as nodes in the molecular graph, we could obtain atom weights by gradient backward propagation and visualized according to Fig. 1 (the darker the color of the node is, the greater the impact on the target is). Based on this, we could identify atomic features that are significant

for the target and then optimize chemical structures for drug design. Taking the multitask model into account, the atoms having relatively high contributions play big roles for all tasks. Here, we believe that these high scoring atoms are crucial for the selectivity of the JAK family. The process of atom weighting calculation could be formulated below:

$$Atom_{m,n} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T \left[W \vec{h}_k || W \vec{h}_j \right] \right)\right)}{\sum_{i \in N_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^T \left[W \vec{h}_k || W \vec{h}_i \right] \right)\right)} \quad (1)$$

This formula expresses the contributions of node j (the neighbor node, a neighbor atom) for node k (the target node, a specific atom) without considering the graph structural information. h_k is the state vector of node k , h_j is the state vector of node j and W is a trainable weight matrix. The mutual attention mechanism $\vec{a}^T (W \vec{h}_k, W \vec{h}_j)$ is applied in the model, parameterized by the weight vector \vec{a} and activated by applying LeakyReLU. The transpose is denoted as T and concatenation is by $||$. The numerator is expressed as the sum of the weights of all neighbor nodes of k .

To determine whether built models have learned the protein binding logic, 38 molecular series were extracted and compiled from Park's works [39]. These molecular series were experimentally synthesized JAK inhibitors and contained activity values for all JAK isoforms. We performed predictions with a trained model for these inhibitors, inspecting and visualizing the atomic weights for fine-tuning analysis. Our goal was to evaluate how faithfully these contributions captured the binding logic for each JAK proteins.

Machine learning approaches

To further enable methodology evaluation, we compared our multitask deep learning model with a state-of-the-art machine learning method called LightGBM [40], which is a distributed gradient boosting framework based on the decision tree algorithm. LightGBM is faster and lower memory consumption in contrast to Extreme Gradient Boosting (XGBoost) [41] that is another gradient boosting framework. It abandons the Level-wise decision tree growth strategy and used a Leaf-wise strategy with a depth limit. Level-wise is an inefficient algorithm because it treats the leaves of the same level indiscriminately, which brings a few unnecessary computational overheads since many leaves indeed have low splitting gain and there is no need to search and split. By contrast, Leaf-wise is more productive, which finds the leaf with the highest splitting gain and then splits it, and so on. Moreover, LightGBM adds a maximum depth limit

on top of Leaf-wise to prevent overfitting while maintaining high efficiency. Given the above considerations, we applied LightGBM algorithm rather than XGBoost. The ECFP4 fingerprint (1024 bits) [42] was chosen for the input of the machine learning model and calculated by PaDEL-descriptor program (version 2.13) [43], which is widely used in QSAR/QSPR tasks to characterize molecules. Furthermore, considering that better performance may be achieved by other types of molecular representations [44], we calculated the molecular descriptors of the compounds using RDKit as a second approach. The model parameters we set to build eight regression models were also based on random search, and the final outcomes were illustrated in Additional file 1: Table S3. In order to proceed model performance comparisons, we guaranteed the consistency of the all datasets (training set, validation set and test set) with deep learning methods.

Model evaluation

The measurements in these data sets are quantitative. We built regression models for the quantitatively measured data sets. The performance of regression models was evaluated by the following metrics: R^2 (the degree of concordance between the predictions and the corresponding observations), MAE (mean absolute error) and RMSE (root-mean-square error). The predictive model could be single-task or multitask. For the multitask models, we calculated the performance metrics for each individual task and reported their average values as global metrics. The formulas are as follows.

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (2)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4)$$

where, y_i presented true values, \hat{y}_i presented predicted values and m is the total number of data sets.

Y-randomization testing

To estimate the impacts of chance correlation, y -randomization was used, which is developed to validate a given regression model and initially proposed by Rucker et al. [45]. In this approach, activity values of four tasks are randomly shuffled, to disrupt the relationship between label values and feature values in the

training data, and construct the model on the basis of unordered data. The procedure is repeated number of times. Ideally, if the new regression models have lower R^2 values for several trials, then the given model is thought to be robust. Inversely, given a regression model, even though the performance of the model is great respect to the training data, if the performance after y -randomization is good as well, then there should be chance correlations in the data set and the model could be overfitted. In the present study, we randomly shuffled the IC_{50} values of training sets and validation sets ten times based on different random seeds. These disordered data were then applied to reconstruct the attentive FP multitask model repetitively, and recorded the R^2 values of the training and validation set for each y -randomization model.

Definition of applicability domain

Defining the applicability domain (AD) of the model is a key component in the five standards of OECD (Organization for Economic Co-operation and Development) on QSAR models, which can be considered as the chemical space of the modeling compound data [46]. Thus, we adopted a methodology based on structural similarity named the Euclidean distance-based method (DM) for AD analysis in this study. The chemical structures were represented by Morgan fingerprints. This method will ultimately obtain a distance threshold (D_T) which can determine whether the compound is within the AD of the model. The detailed formula is as follows:

$$D_T = d_{ave} + Z * \theta \quad (5)$$

where, d_{ave} is the average of the Euclidean distance between each compound and its nearest compound in the training set, θ is the corresponding standard deviation, Z is an optional parameter representing the significance level. First, we calculate the structural similarity between the test set and the training set of compounds by RDKit to get d_{ave} and θ , then keep the k nearest neighbor molecules with the highest similarity as the distance value. If one of these k distances exceeds the threshold of D_T , the compound is considered to be outside of domain (OD) [47]; otherwise, it has fallen into the domain (ID).

Results and discussion

Data analysis

After data processing, 13,898 compounds (including 8087 JAK1, 10,828 JAK2, 4485 JAK3 and 2465 TYK2 inhibitors) were retained to build multitask regression models. Most of the compounds have more than two JAK isoform experimental IC_{50} values ranging from 0.00125 nM to 767,000 nM. The final experimental value distributions

of each JAK isoform dataset were shown in Fig. 2, which was adequate to build a robust activity prediction model.

Chemical diversity analysis

To verify the diversity of collected compounds and rationalization of data set partitioning, we visualized the chemical space with a principal component analysis (PCA) [48]. MACCSkeys generated by RDKit as input of PCA were used to represent JAK ligands. As demonstrated by the chemical space defined by the first three principal components in Fig. 3, a wide distribution of scatters was observed, indicating a high diversity of collected molecules. Besides this, the chemical space of the validation and test sets were completely overlapped with the chemical space of the training set, which implied justifications of data splits. In our model, the training set contained 10,824 molecules, the validation set contained 1529 molecules, and there were 1545 molecules in the test set.

Task relevance analysis

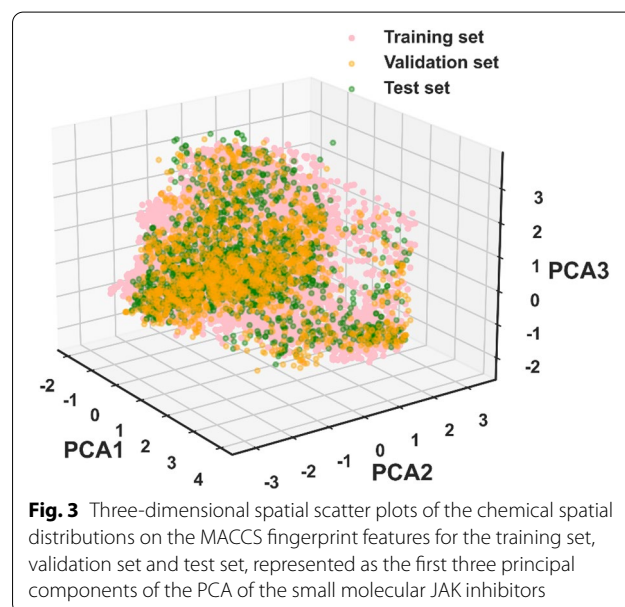
In contrast to transfer learning, multitask learning is more suitable for data sets that have shared molecules with related tasks and can predict different tasks using only one network architecture [49]. To verify that there are some correlations between the individual tasks, we conducted a correlation analysis on the pIC_{50} values of the four tasks, and the findings were shown in Fig. 4a. We also calculated the quantitative estimate of drug-likeness (QED) values for the compounds in the four tasks, and plotted their area under curves in Fig. 4b to estimate the similarity of the compounds, with larger overlap areas reflecting higher compound similarity across the four tasks. QED value is a quantitative metric for assessment of drug-likeness ranging from 0 (all properties unfavorable) to 1 (all properties favorable) [50]. It could be expected that the activity values of all four tasks were



moderately correlated and the four compound datasets were resembled as well.

Optimization of MTATFP model

The prediction performance of our multitask attentive FP (MTATFP) model on the validation dataset was summarized in Additional file 1: Figure S1. In this case, since the size of our overall dataset was relatively small compared to other tasks used for deep learning, and GNN was prone to overfitting [51], we only set two Attentive FP layers and one pooling layer here, and the number of neurons was set to the default value of 300. In the light of the model's parameter search results, we noticed that batch size and dropout had a relatively slight effect on the model, while the learning rate had a significant effect. Models with lower values of learning rate always produce weak predictive power. A large learning rate might cause big prediction fluctuations without learning enough knowledge, while a model with a small learning rate might demand excessive updates to achieve convergence. The best set of hyper-parameters for each category of tasks obtained from the previous random searching process was used to train the most predictive model. Ultimately, we trained models for 217 epochs (according to early stopping state strategy illustrated in Additional file 1: Figure S2) with a batch size of 256 samples, and employed the Adam [52] optimizer with a learning rate of $1e-3$ and a weight decay of $1e-6$. Parameters in the network were updated using MSE Loss which measure mean-squared error as loss functions of the regression tasks. The top performing model had an R^2 value over 0.8



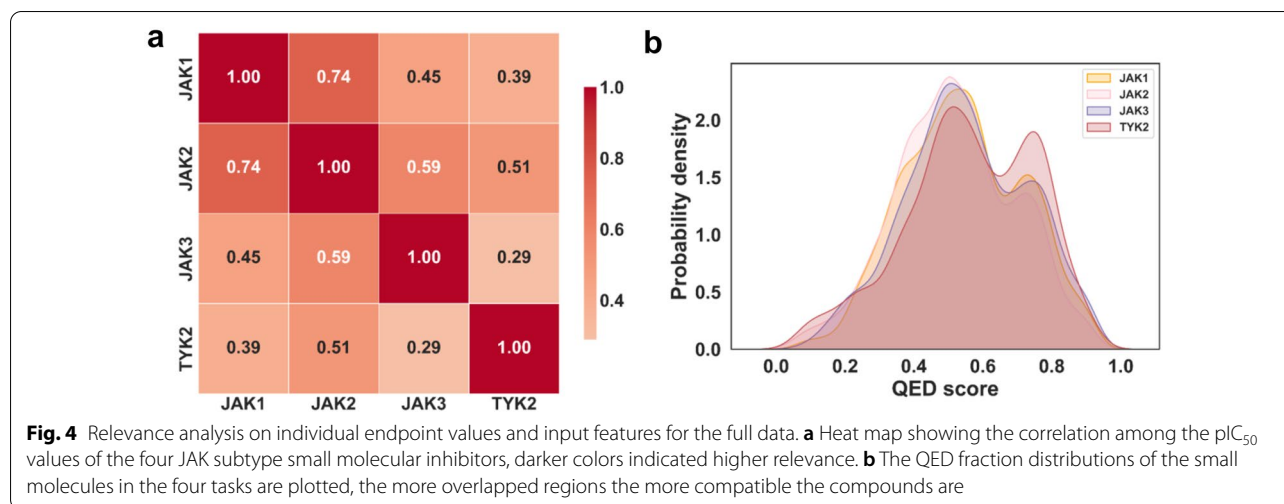


Fig. 4 Relevance analysis on individual endpoint values and input features for the full data. **a** Heat map showing the correlation among the pIC_{50} values of the four JAK subtype small molecular inhibitors, darker colors indicated higher relevance. **b** The QED fraction distributions of the small molecules in the four tasks are plotted, the more overlapped regions the more compatible the compounds are

and the model was selected for the next step in the test set evaluation.

Model performance

MTATFP models showed predictive capabilities according to Fig. 5 and Table 1, with the global R^2 values of 0.96, 0.79 and 0.78 calculated on the training, validation and test sets, respectively. Additionally, the global MAE values of those three datasets (training, validation and test sets) are 0.15, 0.37 and 0.37 (Table 2). The corresponding global RMSE values are 0.23, 0.51 and 0.52 (Table 3). Distributed to every individual task in our model, the R^2 values in training set from 0.20 to 0.27. And in validation set and test set the R^2 values ranged from 0.75 to 0.82, the MAE values ranged from 0.30 to 0.42 and the RMSE values ranged from 0.42 to 0.58. These values indicated that the difficulty of the training task varies. The MTATFP regression model was further validated by the Y-randomization test (Fig. 6), and the global R^2 values of the randomized model was apparently lower than those of the non-randomized model, which demonstrated that the given regression model is robust and not the outcome of chance. Although none of the tasks showed perfect predictive power and the performance on small datasets were worse in some extends, the results are remarkably better than random (proved by Y-Randomization Test), indicating that meaningful molecular graph features related to target endpoints were identified during the learning process.

The tables (Tables 1, 2 and 3) also displayed the assessment results of all single-task models based on Attentive FP (STATFP) networks or LightGBM algorithm. Obviously, studies on four homologous proteins' IC_{50} predictions have shown that multitask learning had great advantages over single-task learning and outperformed

other methods on both validation and test set. In the case of STATFP models that utilized the same framework, the global R^2 for the validation and test set of the MTATFP model improved by 5% and 7%, the global RMSE decreased by 10% and 8%, and the global MAE decreased by 8% and 8%, respectively. Observing the performance of individual tasks in the multitask model, it had significantly improved the model performance on the small data of JAK3 and TYK2, while keeping the model validity constant or even slightly enhancing on the large data set of JAK1 and JAK2 (probably learned the molecular graph knowledge from other tasks), which was in accordance with our anticipations of employing the multitask learning strategy. Versus deep learning methods, the LightGBM method based on molecular descriptors or fingerprints seemed to have close predictive abilities on the small data of JAK3 and TYK2, yet the similar results were not achieved in larger data set. All in all, the MTATFP model yielded the best performance in all validation and test evaluations, producing the most accurate predictions with better generalizability.

Comparison with other studies on JAK selectivity

Recently, Li and colleagues proposed a multitask classification model of 391 kinases including JAK2, JAK3 and TYK2 to distinguish inhibitors from non-inhibitors and achieved good performance [53]. But the classification models did not present accurate bioactivity data and were somewhat insufficient to judge whether a compound was selective or not. For instance, some compounds are selective, where their IC_{50} values for individual kinase isoforms are all below the threshold of 1 μM , but in fact they have a difference multiplicity greater than twice. Or, molecules determined to be selective by the classification model, some of which have IC_{50} values around the threshold, are

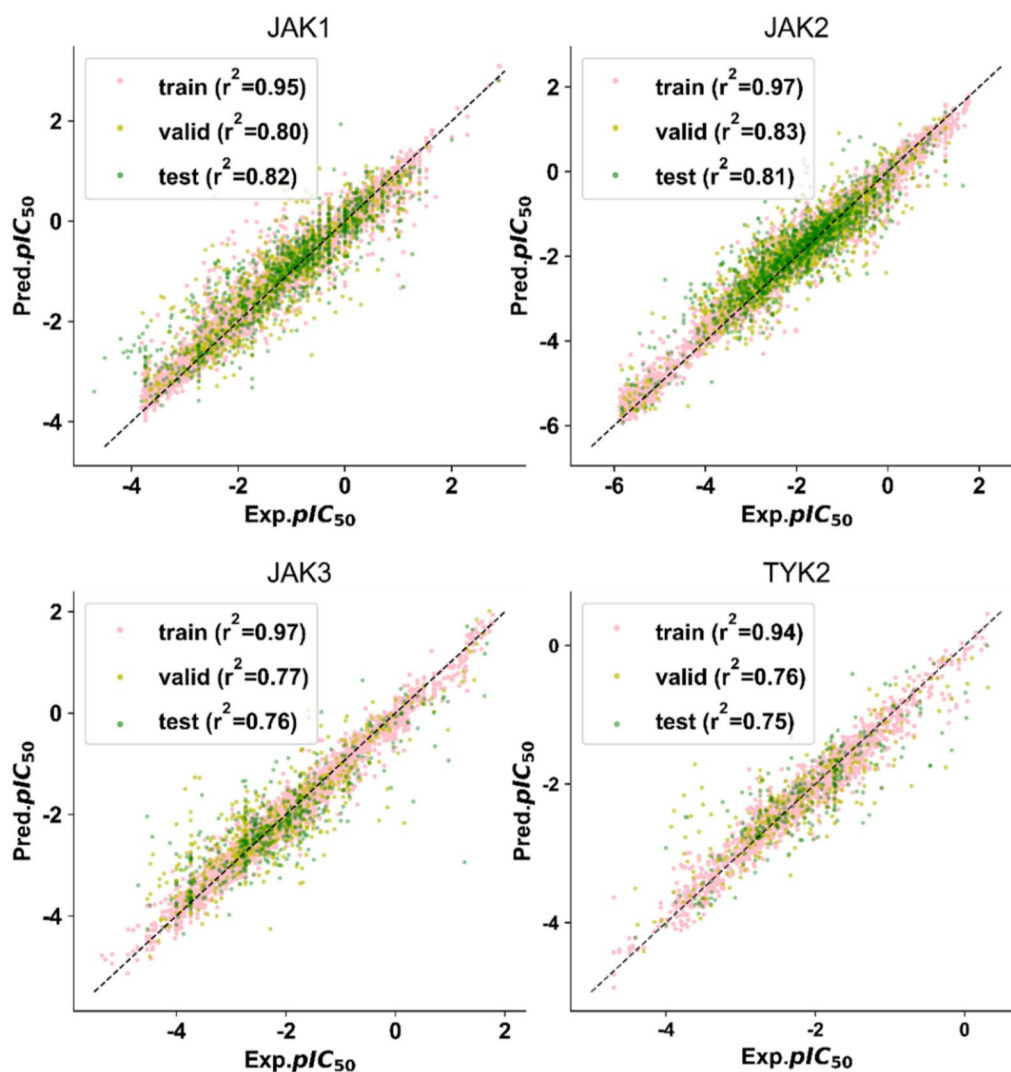


Fig. 5 Linear scatter plot of the training, validation and test set in our MTATFP model. The closer the scatter points are to the straight line, the better the approximation of the model predictions to the true values is

indeed non-selective compounds. To facilitate the comparison with Li's work, we maintained the evaluation metrics and divided the predicted IC_{50} values into active and inactive compounds at $1 \mu M$ to assess their AUC values. For Davis's datasets, a K_d values over $1 \mu M$ was defined as signifying inactivity. All thresholds were set in consistent with the literature and the results of the comparisons were represented in Fig. 7. It could be seen that our MTATFP model had similar performance compared to Li's MTDNN model in the discriminations of inhibitors for the respective JAK members. Besides, we could extraordinarily provide specific bioactivity values IC_{50} on top of that. Although our model performed slightly worse on several data sets, this was likely associated with the

chemical space of the data set with certain compounds occurring outside the AD of the model (shown in Additional file 1: Figure S3). Overall, MTATFP still exhibited impressive performance on these external datasets, and it would be possible to further improve the generalizability of the model by expanding the size of the training dataset and increasing the diversity of compounds.

In addition, there have been a few other ligand-based VS approaches like other machine learning algorithms and three-dimensional quantitative structure-activity relationship (3D-QSAR) analyses used in the discovery of JAK inhibitors as well, however they only consider drug-likeness but not drug-selectivity since these efforts only gave modeling analysis of IC_{50} values for a specific JAK

Table 1 R² performance of each model on the four tasks

	Methods	Training set	Validation set	Test set
Global	MTATFP	0.96	0.79	0.78
	STATFP	0.93	0.75	0.73
	LightGBM_MD	0.93	0.69	0.70
	LightGBM_ECFP4	0.91	0.71	0.70
JAK1	MTATFP	0.95	0.80	0.82
	STATFP	0.95	0.80	0.80
	LightGBM_MD	0.91	0.75	0.72
	LightGBM_ECFP4	0.91	0.75	0.76
JAK2	MTATFP	0.97	0.83	0.81
	STATFP	0.96	0.82	0.8
	LightGBM_MD	0.94	0.70	0.75
	LightGBM_ECFP4	0.92	0.75	0.71
	Xgboost ^a	0.97 ^a	0.80 ^a	0.80 ^a
JAK3	MTATFP	0.97	0.77	0.76
	STATFP	0.91	0.68	0.70
	LightGBM_MD	0.92	0.69	0.70
	LightGBM_ECFP4	0.90	0.69	0.73
TYK2	MTATFP	0.94	0.76	0.75
	STATFP	0.91	0.69	0.63
	LightGBM_MD	0.93	0.61	0.62
	LightGBM_ECFP4	0.91	0.65	0.61

It showed that the performance of various models (deep learning methods based on MTATFP or STATFP strategies and LightGBM-based machine learning approaches) on each task. The closer the R² value was to 1, the better the model performed

^a is the best results in Yang's work. Although each dataset could not be guaranteed to be the identical, our multitasking model has obvious advantages as well

isomer. Yang and colleagues constructed three groups of regression models based on fingerprints and XGBoost methods to acquire highly potent JAK2 kinase inhibitors [54], the results were shown in Table 1. In contrast to their work, our model exhibited identical predictive capability for JAK2 active small molecules and at the parallel it was able to present the IC₅₀ values of compounds for the other three isoforms with selectivity analysis. The most important thing is that our model breaks the black box by visualizing the atomic weights for model interpretations rather than just giving some prediction outputs. And compared with several 3D-QSAR works for JAK selectivity studies [55–58] which collected only tens or hundreds of compounds to model, our model has a wider AD and higher robustness because of the broader chemical space of the four datasets.

Atom visualization and interpretation

In this part of the study, we defined atoms with weight score above the average as key atoms, and molecular fragments with multiple key atoms were considered as

Table 2 MAE performance of each model on the four tasks

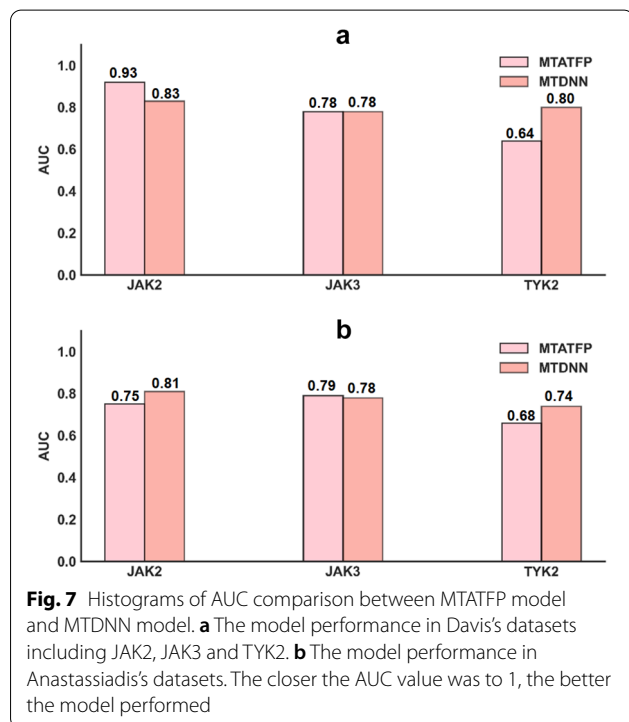
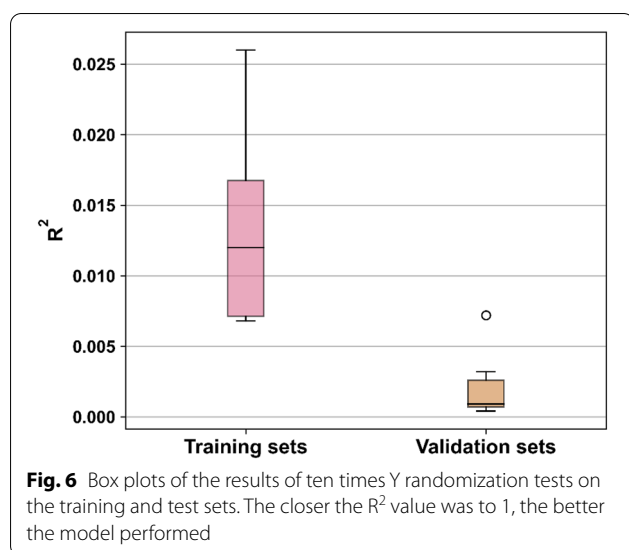
	Methods	Training set	Validation set	Test set
Global	MTATFP	0.15	0.37	0.37
	STATFP	0.21	0.40	0.40
	LightGBM_MD	0.22	0.45	0.44
	LightGBM_ECFP4	0.23	0.43	0.43
JAK1	MTATFP	0.19	0.37	0.39
	STATFP	0.18	0.36	0.39
	LightGBM_MD	0.25	0.45	0.44
	LightGBM_ECFP4	0.24	0.42	0.44
JAK2	MTATFP	0.18	0.38	0.38
	STATFP	0.21	0.40	0.40
	LightGBM_MD	0.22	0.48	0.47
	LightGBM_ECFP4	0.25	0.46	0.46
JAK3	MTATFP	0.16	0.42	0.41
	STATFP	0.27	0.49	0.47
	LightGBM_MD	0.24	0.49	0.48
TYK2	MTATFP	0.15	0.30	0.30
	STATFP	0.19	0.34	0.35
	LightGBM_MD	0.15	0.36	0.38
	LightGBM_ECFP4	0.18	0.37	0.37

The lower the MAE value was, the better the model performed

Table 3 RMSE performance of each model on the four tasks

	Methods	Training set	Validation set	Test set
Global	MTATFP	0.23	0.51	0.52
	STATFP	0.29	0.56	0.57
	LightGBM_MD	0.30	0.62	0.61
	LightGBM_ECFP4	0.33	0.60	0.60
JAK1	MTATFP	0.27	0.52	0.53
	STATFP	0.27	0.51	0.55
	LightGBM_MD	0.34	0.61	0.60
	LightGBM_ECFP4	0.35	0.57	0.60
JAK2	MTATFP	0.26	0.51	0.54
	STATFP	0.28	0.55	0.56
	LightGBM_MD	0.31	0.66	0.62
	LightGBM_ECFP4	0.36	0.63	0.64
JAK3	MTATFP	0.21	0.58	0.58
	STATFP	0.36	0.70	0.66
	LightGBM_MD	0.33	0.67	0.65
	LightGBM_ECFP4	0.38	0.66	0.62
TYK2	MTATFP	0.20	0.44	0.42
	STATFP	0.26	0.49	0.52
	LightGBM_MD	0.21	0.52	0.55
	LightGBM_ECFP4	0.25	0.52	0.52

The lower the RMSE value was, the better the model performed



key substructures. Altering a few key atoms or substructures might have a meaningful impact on the predicted outcomes of four tasks. To estimate whether the atom weightings were statistically significant, we introduced a rank sum test [59]. Two sets of atom weightings were tabulated as A and B. A was the atom weightings in the substructure that we deemed important, and B was the weightings of the remaining atoms in that molecule. In a rank sum test, if the hypothesis that A is much greater

than B is valid, the p-value is less than 0.05 which means the atoms in A are much more important than B. To facilitate the presentation of those findings, the atom weightings of each molecule in JAK inhibitor series were regularized and visualized. Three chosen areas were modified to heighten the selectivity of JAK1 inhibitors in Park's works (shown in Fig. 8), some of which would be interpreted by our trained multitask models. We predicted the IC_{50} values of the lead compound **17a** for the four JAK isoforms and calculated its atom weights. As shown in Fig. 8, the cyclohexane in the B part showed a dark green color, which we assumed to be a key substructure, and performed a rank sum test on the atoms in it and found that the hypothesis held and the average atomic weight (0.63) of this substructure was also remarkably higher than that (0.41) of the remaining atoms, so we determined that the cyclohexyl amine is a key substructure. Although the average atom weights of the A and C parts was a bit lower than the total atom weight (0.46), they were also potential to affect the final outcomes due to the presence of some key atoms, which was consistent with the descriptions in the literature and we could utilize these for drug design.

The JAK1 selective inhibitors were designed by fine-tuning the substituents or scaffolds and observing the deviation of the four IC_{50} values. Methyl amide, cyclopropyl amide and cyclopentyl amides were compared in the A part, and it was found that the amide bonds of all three compounds (**17c**, **17i** and **17k**, shown in Fig. 9) were displayed in dark color. However, the average atom weights of A part decreased as the amide volume became bulkier and the binding affinity of the compounds to JAK changed at the same time, suggesting that amide might be a JAK isomer selective switch especially the methyl amide due to it had the highest atom weight and the greatest influence. Our model predictions were in agreement with experimental results in the literature [39]. We further investigated the effect of the size and shape of the cyclic amine group in the B part on the selectivity of JAK by determining three basic scaffolds and simultaneously ensuring that the other substituents were the same, including 3-aminopyrrolidine (**19c**), piperidin-4-ylmethyl (**18c**) and 4-aminoazepane (**30c**) (presented in Fig. 9). The results showed that the inhibitory activity against JAK1 was slightly increased but selectivity decreased obviously with the size of the cyclic amine group, which also supported that the cyclic amine group was a key substructure in concordance with the literature descriptions [39]. At next stage, we were interested in investigating whether our model could correctly capture the stereochemistry of the compounds and whether the cis-trans isomers of the compounds were correlated with the selectivity of the JAK inhibitors. According to

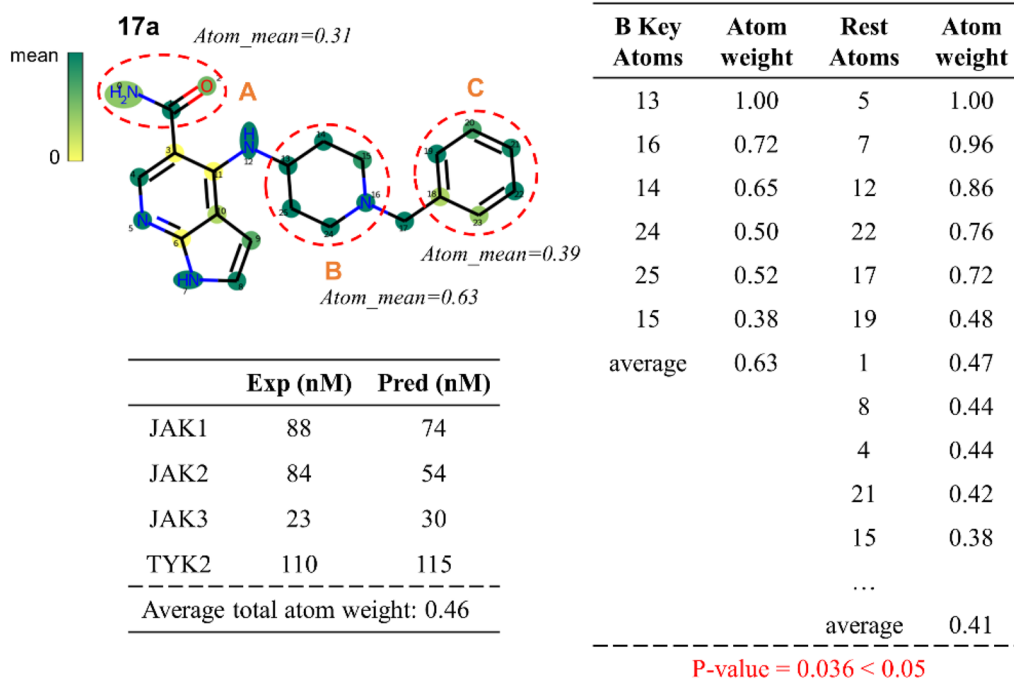


Fig. 8 The atom weight visualization of the lead compound **17a**. Three areas (A, B, C) were chosen to be modified for the selective JAK1 inhibitor design. Atom_mean represented the average of the atom weights in the highlighted region in red and the deeper the color in the graph is, the higher the value of the atom weights is. The table in the figure exhibited some of the atom weight values with their atom indexes and the predicted IC_{50} values

the literature [39], **38a** was more selective for JAK1 than JAK2, JAK3 and TYK2, respectively, while the enantiomer **38b** has a higher IC_{50} value for JAK1 which was far outweighed by **38a**. Under our calculations, although our model had some variations in the predictions for **38b**, the overall tendency was the consistency with the experimental results, and the weights of the atoms near one of the E bonds were changed to some extent, which indicated that our model have learned certain knowledge of chemical space.

Analysis of applicability domain

The AD setting can avoid over-prediction bias arising from the significant characteristic differences between test and training chemicals, which means that it is necessary to evaluate the AD of a model to identify the reliability of the prediction for different molecules [60]. By calculating the distance matrix, we obtained d_{ave} of 0.1262 and θ of 0.2522. Next, in order to find the best AD, we set different k values and Z values, finally received the corresponding threshold D_T and the corresponding number of OD compounds, which were given in Table 4. As observed, the consequent increase in Z values and decrease in k resulted in an accompanying growth of D_T and a constant decline in the compounds outside the AD.

Subsequently, we used our MTATFP model to predict the ID and OD compounds in the test set at different values of k and D_T , and the performance of each data set was shown in Table 5. By comparing and analyzing the results, we noticed that the overall evaluation metrics of the model were all improved when $D_T = 0.06315$ ($k=3$, $Z=-0.25$) (compared to the complete test set without removing the OD compounds) and were able to distinguish to the maximum extent between ID and OD compounds (the prediction performance of ID compounds was significantly better than that of OD compounds). The findings indicated that our defined AD is appropriate for the proposed MTATFP model and could allow the model to serve more accurately in practical applications.

In addition, it is worth to notice that our MTATFP model not only has an excellent prediction ability for ID compounds ($R^2=0.79$, MAE=0.36, RMSE=0.51), but also has a good prediction performance for OD compounds ($R^2=0.72$, MAE=0.48, RMSE=0.63). We believed that this phenomenon could reflect the advantage of a multitask learning strategy, which allows sub-tasks to learn from each other. Even if a compound is missing information from one task, it will still be determined by stealing knowledge from other tasks, which

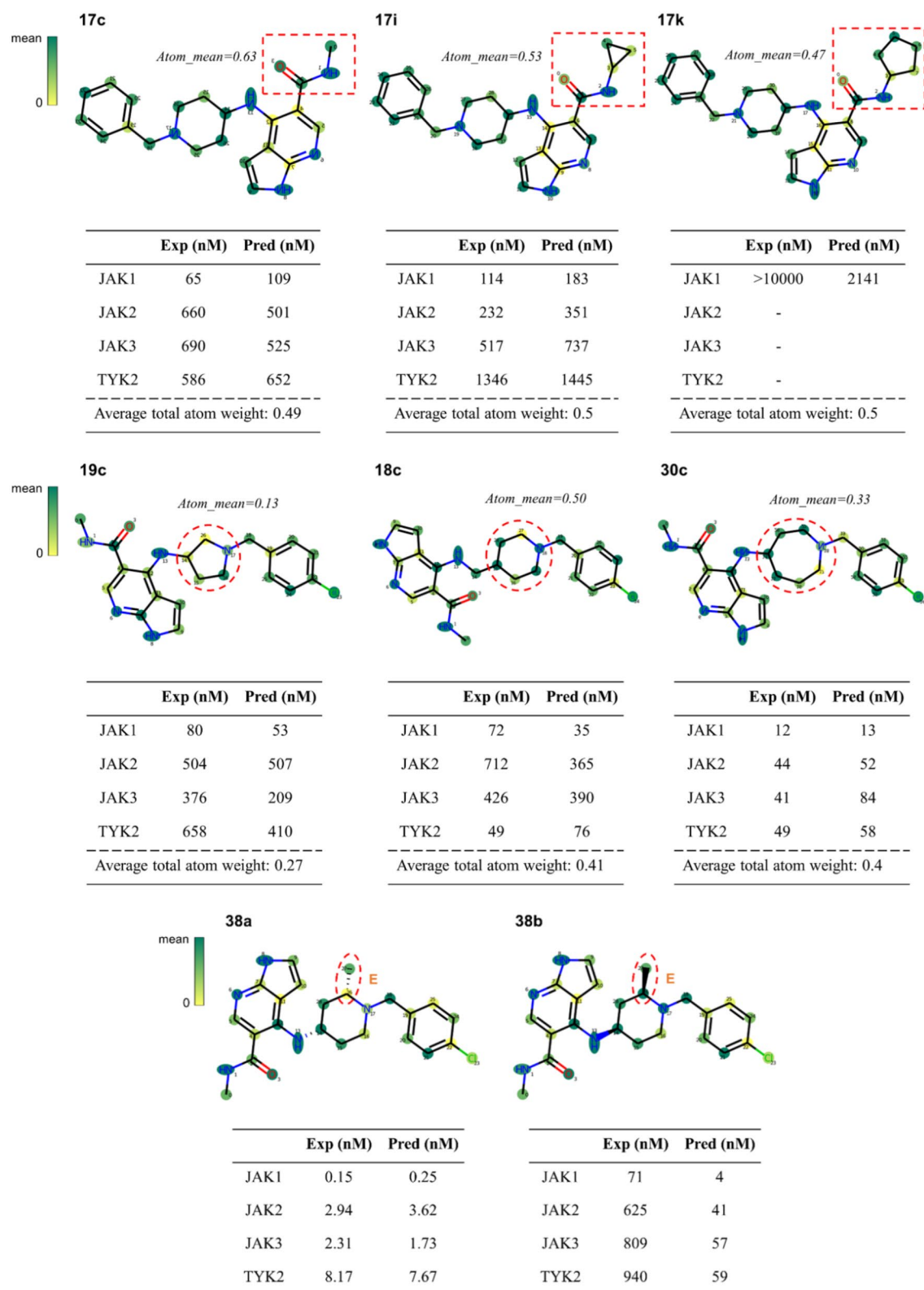


Fig. 9 Several examples of atom weight visualization for JAK inhibitor small molecules

Table 4 The amounts of compounds outside the applicability domain in the test set at different Z and k values

Z	- 0.25	- 0.2	- 0.15	- 0.1
k				
3	130	92	67	43
4	176	117	87	51
5	226	150	114	69
D_T	0.06315	0.07576	0.08837	0.1010

may correspondingly increase the fault tolerance of the model.

Conclusions

Designing selective JAK inhibitors has been a daunting challenge owing to the extremely high homology among individual isoforms. Although traditional QSAR models tend to have perfect predictive power, there is still no

guarantee that the predictions could guide selective drug design. Here, we constructed an MTATFP regression model to predict the pIC_{50} values of small molecules to four JAK isoforms and got atom weights to identify key atoms and substructures important to the target selectivity. Then we used these key substructures to fine-tune the compounds, and further defined the applicability domain of the model. The results indicated that the constructed model could effectively learn the interactions between small molecule ligands, each JAK isoform and the key substructures recognized can correctly guide JAK-selective inhibitor design. The multitask learning strategy also significantly improved the model performance for small data sets, giving an extended error tolerance. Overall, our JAK-selective virtual screening platform offers the advantages of speed, accuracy and interpretability for quantitative prediction of selective JAK inhibitors.

Table 5 The evaluation performance of compounds both inside domain (ID) and outside the domain (OD) in the test set at different Z and k values

	R^2		MAE		RMSE	
	ID	OD	ID	OD	ID	OD
k3 $Z \setminus 0.25$	0.79	0.72	0.36	0.48	0.51	0.63
k4 $Z \setminus 0.25$	0.79	0.76	0.36	0.44	0.51	0.59
k5 $Z \setminus 0.25$	0.79	0.74	0.36	0.44	0.50	0.59
k3 $Z \setminus 0.2$	0.79	0.73	0.37	0.45	0.51	0.59
k4 $Z \setminus 0.2$	0.79	0.74	0.36	0.45	0.51	0.59
k5 $Z \setminus 0.2$	0.79	0.75	0.36	0.44	0.51	0.59
k3 $Z \setminus 0.15$	0.79	0.70	0.37	0.45	0.51	0.57
k4 $Z \setminus 0.15$	0.78	0.76	0.37	0.45	0.51	0.58
k5 $Z \setminus 0.15$	0.79	0.75	0.36	0.44	0.51	0.58
k3 $Z \setminus 0.1$	0.78	0.74	0.37	0.44	0.52	0.57
k4 $Z \setminus 0.1$	0.78	0.77	0.37	0.44	0.52	0.57
k5 $Z \setminus 0.1$	0.79	0.75	0.37	0.45	0.51	0.59

Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; JAK/STAT: Janus kinase/signal transducers and activators of transcription; FDA: Food and drug administration; UC: Ulcerative colitis; ATP: Adenosine triphosphate; VS: Virtual screening; QSAR: Quantitative structure–activity relationship; GNN: Graph neural networks; MTATFP: Multitask attentive FP; GAT: Graph attention network; XGBoost: Extreme gradient boosting; MAE: Mean absolute error; RMSE: Root-mean-square error; AD: Applicability domain; DM: Distance-based method; OD: Outside of domain; ID: Inside of domain; PCA: Principal component analysis; QED: Quantitative estimate of drug-likeness; STATFP: Single-task models based on Attentive FP.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-022-00593-9>.

Additional file 1: Table S1. Approved JAK inhibitors and their current indications. **Table S2.** Some molecular graph features computed with RDKit. **Table S3.** Parameters' settings of LightGBM based model for four tasks. **Figure S1.** Heat map of optimal hyper-parameters search for a MTATFP model. Here, the default parameters were learning rate (0.1, 0.01, 0.001 and 0.0001), drop out (0.2, 0.3, 0.4, 0.5), and search for parameters at different batch size (64, 128, 256) to determine the best ones. The darker the color is, the better the R^2 value of the validation set is. **Figure S2.** A curve graphs of Loss and R^2 during training process on multitask models. Early stopping criterion for training is that the R^2 on validation set is no longer improving in 20 epochs and get the best epoch 217 eventually. **Figure S3.** The chemical spatial distributions of the training, Davis and Anastassiadis dataset. It represented as the first three principal components of the PCA of the JAK small molecular inhibitors.

Acknowledgements

Not applicable.

Authors' contributions

YW designed and performed the research and drafted the manuscript. YW, YG, CL, YG were involved in executing the experiments, ZW and WL provided technical support, TY and GL supervised the study. All authors read and approved the final manuscript.

Funding

The current research was supported by the financial supports from the National Key Research and Development Program of China (Grant 2019YFA0904800), the National Natural Science Foundation of China (Grants 81872800 and 82173746) and Shanghai Frontiers Science Center of Optogenetic Techniques for Cell Metabolism (Shanghai Municipal Education Commission, Grant 2021 Sci & Tech 03-28).

Availability of data and materials

The software used to execute the study is freely available as follows: PaDEL-Descriptor (<http://padel.nus.edu.sg/software/padeldescriptor>). Attentive FP framework was trained with the Deep Graph Library Python (DGL) package (version 0.6.0) [35] with cuda 10.1 and the dgllife extension (github.com/awslabs/dgl-livesci), which ran on the GPU version of the PyTorch framework (version 1.5.0) [37]. Molecular structures were handled and translated into molecular graphs using RDKit (version 2020_03_1) [26]. All the codes, trained models, and datasets are available on our GitHub (<https://github.com/Yimeng-Wang/JAK-MTATFP>).

Declarations**Competing interests**

The authors declare no competing financial interest.

Received: 17 December 2021 Accepted: 26 February 2022

Published online: 15 March 2022

References

- Spinelli FR, Colbert RA, Gadina M (2021) JAK1: number one in the family; number one in inflammation. *Rheumatology (Oxford)* 60(Supplement_2):ii3–ii10
- Tisoncik JR, Korth MJ, Simmons CP et al (2012) Into the eye of the cytokine storm. *Microbiol Mol Biol Rev* 76(1):16–32
- O'Shea JJ, Holland SM, Staudt LM (2013) JAKs and STATs in immunity, immunodeficiency, and cancer. *N Engl J Med* 368(2):161–170
- Mehta P, McAuley DF, Brown M et al (2020) COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet* 395(10229):1033–1034
- Spinelli FR, Meylan F, O'Shea JJ, Gadina M (2021) JAK inhibitors: ten years after. *Eur J Immunol* 51(7):1615–1627
- Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223):497–506
- Yang Y, Shen C, Li J et al (2020) Exuberant elevation of IP-10, MCP-3 and IL-1ra during SARS-CoV-2 infection is associated with disease severity and fatal outcome. *J Allergy Clin Immunol* 146(1):119–127
- Leonard WJ, O'Shea JJ (1998) JAKs and STATs: biological implications. *Annu Rev Immunol* 16(1):293–322
- Shuai K, Liu BJNRI (2003) Regulation of JAK–STAT signalling in the immune system. *Nat Rev Immunol* 3(11):900–911
- Rawlings JS, Rosler KM, Harrison DA (2004) The JAK/STAT signaling pathway. *J Cell Sci* 117(8):1281–1283
- De Vries LCS, Wildenberg ME, De Jonge WJ et al (2017) The future of janus kinase inhibitors in inflammatory bowel disease. *J Crohns Colitis* 11(7):885–893
- Leroy E, Constantinescu SN (2017) Rethinking JAK2 inhibition: towards novel strategies of more specific and versatile janus kinase inhibition. *Leukemia* 31(5):1023–1038
- Actis GC, Pellicano R, Fagoonee S et al (2019) History of inflammatory bowel diseases. *J Clin Med* 8(11):1970
- Choy EHS, Miceli-Richard C, Gonzalez-Gay MA et al (2019) The effect of JAK1/JAK2 inhibition in rheumatoid arthritis: efficacy and safety of baricitinib. *Clin Exp Rheumatol* 37(4):694–704
- McCloskey K, Taly A, Monti F et al (2019) Using attribution to decode binding mechanism in neural network models for chemistry. *Proc Natl Acad Sci USA* 116(24):11624–11629
- Shoichet BK (2004) Virtual screening of chemical libraries. *Nature* 432(7019):862–865
- Banerjee S, Biehl A, Gadina M et al (2017) JAK-STAT signaling as a target for inflammatory and autoimmune diseases: current and future prospects. *Drugs* 77(5):521–546
- Bryan MC, Rajapaksa NS (2018) Kinase inhibitors for the treatment of immunological disorders: recent advances. *J Med Chem* 61(20):9030–9058
- Bajusz D, Ferenczy GG, Keseru GM (2016) Discovery of subtype selective janus kinase (JAK) inhibitors by structure-based virtual screening. *J Chem Inf Model* 56(1):234–247
- Watanabe N, Ohnuki Y, Sakakibara Y (2021) Deep learning integration of molecular and interactome data for protein–compound interaction prediction. *J Cheminform* 13(1):1–12
- Gutierrez G (2020) Artificial intelligence in the intensive care unit. *Crit Care* 24(1):101–101
- Zhou J, Cui G, Hu S et al (2020) Graph neural networks: a review of methods and applications. *AI Open* 1:57–81
- Kim S, Chen J, Cheng T et al (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 47(D1):D1102–D1109
- Mendez D, Gaulton A, Bento AP et al (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 47(D1):D930–D940
- Gilson MK, Liu T, Baitaluk M et al (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res* 44(D1):D1045–D1053
- Landrum G, Tosco P, Kelley B (2020) rdkit/rdkit: 2020_03_1 (Q1 2020) Release 10
- Davis MI, Hunt JP, Herrgard S et al (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 29(11):1046–1051
- Anastassiadis T, Deacon SW, Devarajan K et al (2011) Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* 29(11):1039–1045
- Sutherland JJ, Gao C, Cahya S et al (2013) What general conclusions can we draw from kinase profiling data sets? *Biochim Biophys Acta Proteins Proteom* 1834(7):1425–1433
- Veličković P, Cucurull G, Casanova A, et al (2017) Graph attention networks. *ArXiv Preprint arXiv:1710.10903*
- 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(16):8749–8760
- Xu B, Wang N, Chen T, et al (2015) Empirical evaluation of rectified activations in convolutional network. *ArXiv Preprint arXiv:1505.00853*
- Xu Y, Ma J, Liaw A et al (2017) Demystifying multitask deep neural networks for quantitative structure–activity relationships. *J Chem Inf Model* 57(10):2490–2504
- Ruder S (2017) An overview of multi-task learning in deep neural networks. *ArXiv Preprint arXiv:1706.05098*
- Wang M, Zheng D, Ye Z, et al (2019) Deep graph library: a graph-centric, highly-performant package for graph neural networks. *ArXiv Preprint arXiv:1909.01315*
- Li M, Zhou J, Hu J et al (2021) Dgl-lifesci: an open-source toolkit for deep learning on graphs in life science. *ACS Omega* 6(41):27233–27238
- Paszke A, Gross S, Massa F et al (2019) Advances in neural information processing systems. *Adv Neural Inf Process Syst* 32:8024–8035
- Vellido A, Martín-Guerrero JD, Lisboa PJ (2012) Making machine learning models interpretable. *ESANN* 12:163–172
- Park E, Lee SJ, Moon H et al (2021) Discovery and biological evaluation of N-Methyl-pyrrolo[2,3-b]pyridine-5-carboxamide derivatives as JAK1-selective inhibitors. *J Med Chem* 64(2):958–979
- Ke G, Meng Q, Finley T et al (2017) Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 30:3146–3154

41. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp 785–794
42. Hall LH, Kier LB (1995) Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J Chem Inf Comput Sci* 35(6):1039–1045
43. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32(7):1466–1474
44. Jiang D, Wu Z, Hsieh CY 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 Cheminform* 13(1):1–23
45. Rücker C, Rücker G, Meringer M (2007) Y-Randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 47(6):2345–2357
46. Bhatia S, Schultz T, Roberts D et al (2015) Comparison of cramer classification between toxtree, the OECD QSAR Toolbox and expert judgment. *Regul Toxicol Pharmacol* 71(1):52–62
47. Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* 13(34):3494–3504
48. Ma S, Dai Y (2011) Principal component analysis based methods in bioinformatics studies. *Brief Bioinform* 12(6):714–722
49. Li X, Fourches D (2020) Inductive transfer learning for molecular activity prediction: Next-Gen QSAR Models with MolPMoFIT. *J Cheminform* 12(1):1–15
50. Bickerton GR, Paolini GV, Besnard J et al (2012) Quantifying the chemical beauty of drugs. *Nat Chem* 4(2):90–98
51. Rong Y, Huang W, Xu T, et al (2019) Dropedge: towards deep graph convolutional networks on node classification. *ArXiv Preprint arXiv:1907.10903*
52. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *ArXiv Preprint arXiv:1412.6980*
53. Li X, Li Z, Wu X et al (2019) Deep learning enhancing kinome-wide polyparmacology profiling: model construction and experiment validation. *J Med Chem* 63(16):8723–8737
54. Yang M, Tao B, Chen C et al (2019) Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of JAK2 inhibitors. *J Chem Inf Model* 59(12):5002–5012
55. Zhu J, Yu Q, Cai Y et al (2020) Theoretical exploring selective-binding mechanisms of JAK3 by 3D-QSAR, molecular dynamics simulation and free energy calculation. *Front Mol Biosci* 7:83
56. Vrontaki E, Melagraki G, Afantitis A et al (2017) Searching for novel Janus kinase-2 inhibitors using a combination of pharmacophore modeling, 3D-QSAR studies and virtual screening. *Mini Rev Med Chem* 17(3):268–294
57. Balupuri A, Balasubramanian PK, Cho SJ (2020) 3D-QSAR, docking, molecular dynamics simulation and free energy calculation studies of some pyrimidine derivatives as novel JAK3 inhibitors. *Arabian J Chem* 13(1):1052–1078
58. Itteboina R, Ballu S, Sivan SK et al (2016) Molecular docking, 3D QSAR and dynamics simulation studies of imidazo-pyrrolopyridines as janus kinase 1 (JAK 1) inhibitors. *Comput Biol Chem* 64:33–46
59. Steel RG (1960) A rank sum test for comparing all pairs of treatments. *Technometrics* 2(2):197–207
60. Horvath D, Marcou G, Varnek A (2010) A unified approach to the applicability domain problem of QSAR models. *J Cheminform* 2(1):1–1

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

