Chinese Journal of Chemical Engineering 84 (2025) 201–210

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Prediction of ionic liquid toxicity by interpretable machine learning

Haijun Feng [1,*], Li Jiajia [2], Zhou Jian [2,*]

[1] School of Computer and Software, Shenzhen Institute of Information Technology, Shenzhen 518172, China
[2] School of Chemistry and Chemical Engineering, Guangdong Provincial Key Laboratory for Green Chemical Product Technology, South China University of Technology, Guangzhou 510640, China

ARTICLE INFO

Article history:
Received 15 January 2025
Received in revised form
8 March 2025
Accepted 23 April 2025
Available online 7 June 2025

Keywords:
Ionic liquids
Toxicity
Machine learning
Model
Prediction
Interpretation

ABSTRACT

The potential toxicity of ionic liquids (ILs) affects their applications; how to control the toxicity is one of the key issues in their applications. To understand its toxicity structure relationship and promote its greener application, six different machine learning algorithms, including Bagging, Adaptive Boosting (AdaBoost), Gradient Boosting (GBoost), Stacking, Voting and Categorical Boosting (CatBoost), are established to model the toxicity of ILs on four distinct datasets including Leukemia rat cell line IPC-81 (IPC-81), Acetylcholinesterase (AChE), *Escherichia coli* (*E.coli*) and *Vibrio fischeri*. Molecular descriptors obtained from the simplified molecular input line entry system (SMILES) are used to characterize ILs. All models are assessed by the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient (R^2). Additionally, an interpretation model based on SHapley Additive exPlanations (SHAP) is built to determine the positive and negative effects of each molecular feature on toxicity. With additional parameters and complexity, the Catboost model outperforms the other models, making it a more reliable model for ILs' toxicity prediction. The results of the model's interpretation indicate that the most significant positive features, SMR_VSA5, PEOE_VSA8, Kappa2, PEOE_VSA6, SMR_VSA5, PEOE_VSA6 and EState_VSA1, can increase the toxicity of ILs as their levels rise, while the most significant negative features, VSA_EState7, EState_VSA8, PEOE_VSA9 and FpDensity-Morgan1, can decrease the toxicity as their levels rise. Also, an IL's toxicity will grow as its average molecular weight and number of pyridine rings increase, whereas its toxicity will decrease as its hydrogen bond acceptors increase. This finding offers a theoretical foundation for rapid screening and synthesis of environmentally-benign ILs.

1. Introduction

Ionic liquids (ILs) are novel liquid compounds composed of specific cations and anions, they exist as liquids at or near room temperature. They possess high electrical conductivity, high thermal stability, low volatility, and they are non-polluting and easily recyclable. ILs are extensively employed in the fields of energy, separation, electrochemistry, greenhouse gas capture, catalysis, medicines and other fields [1–3]. Although ILs have many excellent properties, they are not completely non-toxic. Research indicates that, some ILs are difficult to biodegrade; they may accumulate in aquatic environments, posing potential hazards to both the ecological environment and human health. IL 1-octyl-3-methylimidazolium was found in the soil around a landfill in Northeast England, which could potentially trigger primary biliary cholangitis [4]. Therefore, more attentions should be paid on the safety and toxicity of ILs.

With the widespread applications of ILs, the range of their toxic effects will also increase. How to control the toxicity of ILs is an urgent issue to be addressed in their development and utilization. The toxicity of ILs depends on the structure and composition of the cations and anions. Over the past few years, numerous researchers have investigated the toxicity of ILs. Jakubowska *et al.* [5] tested the toxicity of five quaternary ammonium ILs on *Vibrio fischeri* and other organisms, they found that the toxicity was dependent upon the quantity of cations present in the molecule: the more cations, the higher the toxicity. Jeremias *et al.* [6] studied the biotoxicity of imidazolium and cholinium ILs on large water fleas through RNA sequencing, their results indicated that the alkyl chain is a driving

This article is part of a special issue entitled: AI for Chemical Engineering published in Chinese Journal of Chemical Engineering.

* Corresponding authors.
E-mail addresses: fenghj@sziit.edu.cn (H. Feng), jianzhou@scut.edu.cn (Z. Jian).

https://doi.org/10.1016/j.cjche.2025.04.018

factor for the toxicity of ILs. Zhang *et al.* [7] used high-performance liquid chromatography to study the acute toxicity of 1-hexyl-3-methylimidazolium and 1-alkyl-3-methylimidazolium nitrate ILs on zebrafish. They found that the most toxic was $[C_{12}mim]NO_3$, and the least toxic was $[C_2mim]NO_3$, which indicates that ILs with longer alkyl chain lengths have greater acute toxicity impact on aquatic organisms. Although experiments can somewhat analyze the toxicity of ILs, the variety of ILs is vast, with up to 1 trillion types [8]. Using the Microtox Toxicity Test kit, Villanueva *et al.* [9] evaluated the acute toxicity of six ammonium ILs with increasing alkyl chain length ($n = 2, 3, 4, 5, 6, 8$). They discovered that the toxicity of ILs rose with the length of the alkyl chain. Relying solely on experimental methods to explore the potential toxicity of ILs is far from sufficient. Machine learning, with its efficient and accurate predictive capabilities, has been applied in recent years to the study of ILs' melting points [10,11], ILs' capture of acidic gases [12–17] and other areas [18–20]. Machine learning can also effectively analyze and study the potential toxicity of ILs [21]. Fan *et al.* [22] modeled the toxicity of ILs to *Vibrio fischeri* using two machine learning techniques: Random Forest and eXtreme Gradient Boosting. It was shown that the machine learning approach was a useful tool for forecasting and completing the toxicity data for ILs. Zhao *et al.* [23] established a comprehensive database of ILs' toxicity containing over 4000 data points. Based on this database, they predicted the toxicity (EC_{50} values) of several ILs using support vector machine (SVM) and multiple linear regression (MLR) on the IPC-81 leukemia rat cell line. The results showed that the SVM model is more reliable for predicting ILs' toxicity and that raising the relative number of oxygen atoms in the molecule can decrease the toxicity of ILs. Sosnowska *et al.* [24] used 1 global and 6 local quantitative structure-activity relationship models to predict the toxicity of the IPC-81 leukemia rat cell line. They concluded that local models can be used to address specific issues of a set of ILs, but for forecasting the toxicity of a newly synthesized IL, a global model is recommended. Cao *et al.* [25] used three models—MLR, SVM and extreme learning machine (ELM)—to assess ILs' toxicity to the rat leukemia cell line IPC-81. They used quantum chemistry to optimize the structures of 21 anions and 57 cations, and they ultimately found that the ELM model outperformed the other two models in the prediction of ILs toxicity. Ma *et al.* [26] used least squares support vector machine and genetic function approximation algorithms to forecast the cytotoxicity of ILs composed of 9 cations and 44 anions on rat cells (IPC81). They found that both models can be applied to the estimation of cytotoxicity, and that cations were the main factors influencing cytotoxicity due to their hydrophobicity and spatial structure. Yan *et al.* [27,28] conducted a systematic analysis of the inhibition of acetylcholinesterase by 153 ILs utilizing molecular simulation and machine learning. They concluded that combining various machine learning methods could yield more reliable and stable QSAR models, which gave aid to the design of the next generation of biosafe ILs. At the same time, the research group also established a database on the toxicity of ILs, which includes 1183 ILs and over 6700 different biotoxicity data points. Moreover, based on this database, different machine learning models [27,28] have been developed to examine the connection between ILs' toxicity and their structure. This database provides a data foundation for machine learning to accelerate the research on the toxicity of ILs.

Although there are some works applying machine learning methods to the study of ILs' toxicity, research on the potential toxicity to different organisms is still insufficient, and there are two main issues. First, ILs are diverse, and their toxicity varies across different organisms. Current studies mostly focus on the toxicity of the IPC-81 leukemia rat cell line, with less research on other systems. Second, the current machine learning models still focus on toxicity prediction. Due to the limitations of traditional machine learning, they lack the ability to explain the result and analyze the molecular structures that influence the toxicity of ILs. SHapley Additive exPlanations (SHAP) [29] is one of the most promising tools for explaining any machine learning model's output; it is able to measure each feature's contribution to the model's prediction. SHAP values can be used to evaluate which features are most crucial to the model and how they impact the results. Based on a game-theoretic method, SHAP [29] connects local explanations with the optimal credit distribution by utilizing the classic SHapley values from game theory and their extensions.

In this work, six different machine learning models, including Bagging [30], Adaptive Boosting (AdaBoost) [31], Gradient Boosting (GBoost) [32], Stacking [33], Voting and Categorical Boosting (CatBoost) [34,35], are constructed to forecast the toxicity of ILs on four living organisms including Leukemia rat cell line IPC-81 (IPC-81), Acetylcholinesterase (AChE), *Escherichia coli* (*E.coli*) and *Vibrio fischeri*. We initially access each model's performance on four different datasets. Further, we utilize the explanation model SHAP to interpret the results and analyze the contributions of each feature in the intelligent models. The outcomes of this work can be used for the rapid screening of ILs' toxicity and accelerate the development of ILs that are safe for the environment and biological systems.

2. Models and Methodology

2.1. Machine learning models

Six different machine learning techniques are employed to determine the best way to estimate ILs' toxicity: Bagging [30], AdaBoost [31], GBoost [32], Stacking [33], Voting and CatBoost [34,35]. These algorithms are ensemble methods, which can balance out each model's weakness and build a more accurate model during training.

Bagging [30] is an ensemble approach that fits base regressors and creates a final prediction based on voting from its sub regressor. The Decision Tree is selected as the base_estimator in this study. There are 1000 base estimators, and the maximum depth of each estimator is 5.

AdaBoost [31] reweights the training samples in an effort to transform weak learners into strong learners. The Decision Tree is likewise selected as the base_estimator in this study, and there are 1000 base estimators. The learning rate is 1.0; each regressor's contribution rises with a higher learning rate.

GBoost [32], another boosting ensemble technique, enhances the model by combining the Gradient Descent and Boosting algorithms. It corrects the gradients of current weak learners in each boosting step. In this study, the maximum depth of each regression estimator is 5, the number of boosting stages is set to 1000 and the learning rate is 0.1.

The stacking [33] approach is another ensemble technique that uses the output of several regressors as input to a final regressor, aiming at obtaining a more accurate model. It exploits the power of each regressor. In this study, the basic regressors are trained using Bagging [30], AdaBoost [31] and GBoost [32], whereas the final estimator is the Random Forest algorithm [36,37].

The voting algorithm is an ensemble algorithm that averages the individual predictions from a variety of machine learning techniques to provide a final prediction. Four popular algorithms—Bagging [30], AdaBoost [31], GBoost [32] and Random Forest [36,37]—were combined in this study's voting method. To balance out each algorithm's shortcomings and produce a more accurate result, uniform weights were applied to the four algorithms, which was ultimately decided by a simple majority vote.

CatBoost [34,35] is a sophisticated boosting package created especially to solve the difficulties associated with managing

categorical data in machine learning. Due to its capability to generate high-performance models without requiring extensive data preparation, CatBoost [34,35] is an open-source technique that has gained a lot of popularity very rapidly. Because CatBoost [34,35] can handle both numerical and categorical input characteristics with initially ordered goal statistics, it is a better choice for jobs requiring complex, real-world datasets. By including random permutation into the ordered boosting, it can successfully avoid overfitting. In this study, the maximum number of trees is 300 and the learning rate is 0.1.

2.2. Interpretability model of ILs' toxicity

After thorough parameter tuning, the machine learning model can typically produce accurate results, but it still remains "black boxes". How it makes these predictions is beyond our comprehension. In order to understand the results of the model, SHAP [29] is introduced and the interpretation scheme proposed in this work is illustrated in Fig. 1. The inputs are the molecular features of each IL; the model will use these features to forecast an IL's toxicity value. For each feature in the model, a SHAP value is computed and allocated; features with positive SHAP values have a positive influence on the prediction, while those with negative values have a negative effect. The larger the absolute value of the SHAP value of a feature, the greater its impact. From this model interpretability, we can understand how each feature contributes to the toxicity of an IL.

2.3. ILs' toxicity datasets

In this work, four datasets are employed to investigate the toxicity of different ILs. Dataset 1 contains 242 ILs, with toxicity values represented by the half-maximal effective concentration (EC_{50}) of IPC-81 after 72 h. Dataset 2 contains 153 ILs, with toxicity values represented by the EC_{50} of AChE after 5 min. Dataset 3 contains 125 ILs, with toxicity values represented by the minimum inhibitory concentration (MIC) of *E.coli* after 24 h. Dataset 4 contains 165 ILs, with toxicity values represented by the EC_{50} of *Vibrio fischeri* after 30 min. The datasets are derived from the ILs' toxicity database established and made public by Yan *et al.* [27,28]. The ILs in the datasets are characterized by standard SMILES structures, which are the input of the model, and all toxicity values are expressed as their lg values, which are the output of the model.

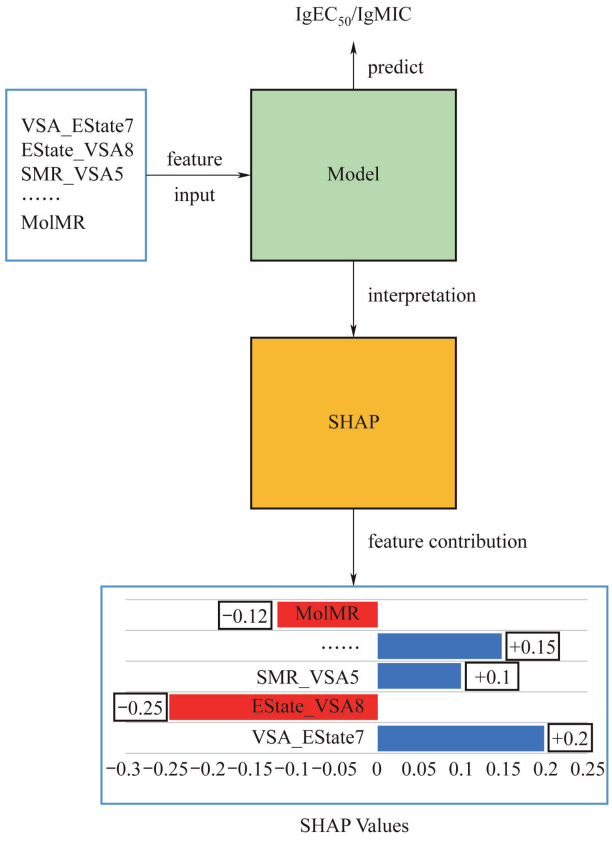

Fig. 1. The interpretation scheme of ILs' toxicity in this work.

The four datasets used in this work are trained individually, with each dataset divided in an 80:20 ratio between training and test sets. A random splitting strategy is employed in the splitting procedure. Following the separation of the datasets, the model is trained on the training set and its accuracy is assessed on the test set. All of the models are evaluated by a 5-fold cross-validation.

2.4. Molecular descriptors from SMILES

In this work, molecular descriptors [38] are mathematical representations of molecules that encapsulate their physical, chemical, and biological features. Molecular descriptors were produced from SMILES structure, which is a widely used technique for molecular structure representation. The SMILES structures were transformed into molecular descriptors by the module "molecular descriptors" of RDKit [39], which is an open-source toolkit for cheminformatics. To examine the ILs' toxicity prediction, 207 specific features were calculated, including properties, composition, charge, topological descriptors, and so on. These features are defined based on how each atom in the molecule contributes to a molecular property [38]. Specifically, features such as "NumValenceElectrons", "FractionCSP3″, "HeavyAtomCount", "NumAromaticRings", "fr_Al_OH" represent for "the number of valence electrons in the molecule", "the fraction of C atoms that are SP3 hybridized", "the number of heavy atoms in the molecule", "the number of aromatic rings in the molecule" and "the number of aliphatic hydroxyl groups", respectively. Detailed descriptions of all features in this work can be found online [40].

2.5. Metrics

The mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient (R^2) are utilized to assess the models' accuracy. The following equations are the metrics' formulations:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(t_i - p_i)^2 \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - p_i)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}|t_i - p_i| \tag{3}$$

$$R^2 = \frac{\sum_{i=1}^{n}(t_i - \bar{t})^2 - \sum_{i=1}^{n}(t_i - p_i)^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2} \tag{4}$$

where n represents the number of data points; t_i represents the experimental toxicity value of ILs; p_i is the toxicity value predicted by the model; $\bar{t}$ denotes the average laboratory value of toxicity; i represents each IL molecule.

2.6. Modeling methodology

Fig. 2 illustrates the general model implementation methodology used in this work. The original four datasets of ILs all use simplified molecular-input line-entry system (SMILES) structures. First, ILs are converted from SMILES structures to molecular descriptors to obtain the molecular features of the ILs. These features are used as input, and the toxicity values are used as output. Six machine learning algorithms (Bagging [30], AdaBoost [31], GBoost [32], Stacking [33], Voting and CatBoost [34,35]) are employed to train the machine learning model. The trained model is evaluated using MSE, RMSE, MAE and R^2 to obtain the optimal model. Afterwards, the toxicity interpretation model is used to explain the results, obtaining the importance of each molecular feature on toxicity; their positive and negative contributions are also ranked. The IL structure is converted by using RDKit [39]; the code is written in Python, and machine learning is based on the SciKit-Learn [41,42] package. SHAP [29] is utilized to interpret the model.

3. Results and Discussion

3.1. Learning curve of models

The learning curves of four models on four distinct datasets are presented in Fig. 3 to show if the training is overfitting or underfitting. Performance is scored using MSE by the CatBoost [34,35] algorithm. For the model on datasets IPC-81, MSE values plateau with a lower MSE following 140 training examples, suggesting that at least 140 ILs' toxicity data points are necessary for the CatBoost model to perform at its optimal level. For the model on datasets AChE, *E.coli* and *Vibrio fischeri*, more datasets can be better for improving the performance of models. Overall, we can see from the results that the training sets' performance scores eventually converged, indicating that there was neither significant variation nor strong bias during training.

3.2. Prediction of ILs' toxicity

The evaluation results of the 6 models on 4 different datasets (test sets) are shown in Table 1. In general, all six models can have good predictive capabilities on the four different datasets. With the lowest values of MSE, RMSE, MAE and the highest value of R^2, the CatBoost [34,35] model has superior performance among all models. On dataset IPC-81, the CatBoost model's values for MSE, RMSE, MAE and R^2 are 0.230, 0.479, 0.357 and 0.845 accordingly, indicating that the model is reliable. On datasets AChE and E.coli, the CatBoost model also has reliable performance. On the dataset *Vibrio fischeri*, values of MSE, RMSE, MAE and R^2 are 0.440, 0.655, 0.470 and 0.774 accordingly, which indicates the CatBoost model has to be enhanced for this dataset. Meanwhile, the other five models also have poor performance on this dataset; the fact that the *Vibrio fischeri* dataset is tiny and sparse might be the cause. After all, the CatBoost model exhibits optimal behaviors in contrast to the other five machine learning models. With additional parameters and complexity, the CatBoost model is better equipped to manage the complexities and nonlinear connections of the dataset, allowing it to identify more intricate patterns. So, the CatBoost model is chosen for further study.

The parity plot of the experimental toxicity values and predicted ones of the CatBoost model on the test sets of the four datasets is shown in Fig. 4. From Fig. 4, it is evident that the four datasets' predicted values are consistent with experimental values, demonstrating the model's excellent accuracy and reliability.

3.3. Comparison of model performances

As can be seen in Table 2, which compares the performance of model in this work with the models that have been reported, all the reported models, including support vector machine (SVM) [43,44], random forest (RF) [36,37,44], Multilayer Perceptro (MLP) [44,45], ExtremeGradient Boosting (XGBoost) [27,46], artificial neural network (ANN) [27,47], are trained on the same datasets; RMSE, MAE and R^2 values are utilized as the performance metrics. On datasets AChE, MLP [44,45] and XGBoost [46] models perform better than the CatBoost [34,35] model. MLP models have a lower MAE value of 0.156 and a higher R^2 value of 0.825. While the XGBoost model's RMSE value is 0.200 and its R^2 value is 0.840, which is also better than the CatBoost model does. Yet, the CatBoost model has superior performance over the other investigated models on datasets IPC-81, *E. coli* and *Vibrio fischeri*. It has a lower RMSE, MAE and a higher R^2 value. Due to the small dataset, the model's performance on different datasets shows some fluctuations, while the CatBoost model's precision and reliability are generally excellent.

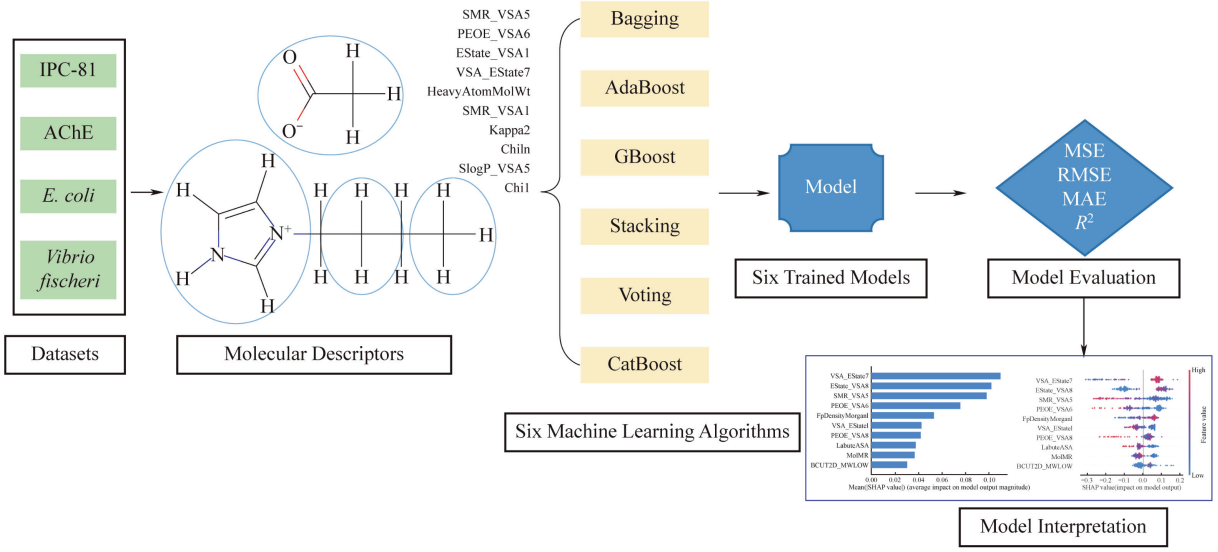

Fig. 2. The overall methodology for modeling used in this work.

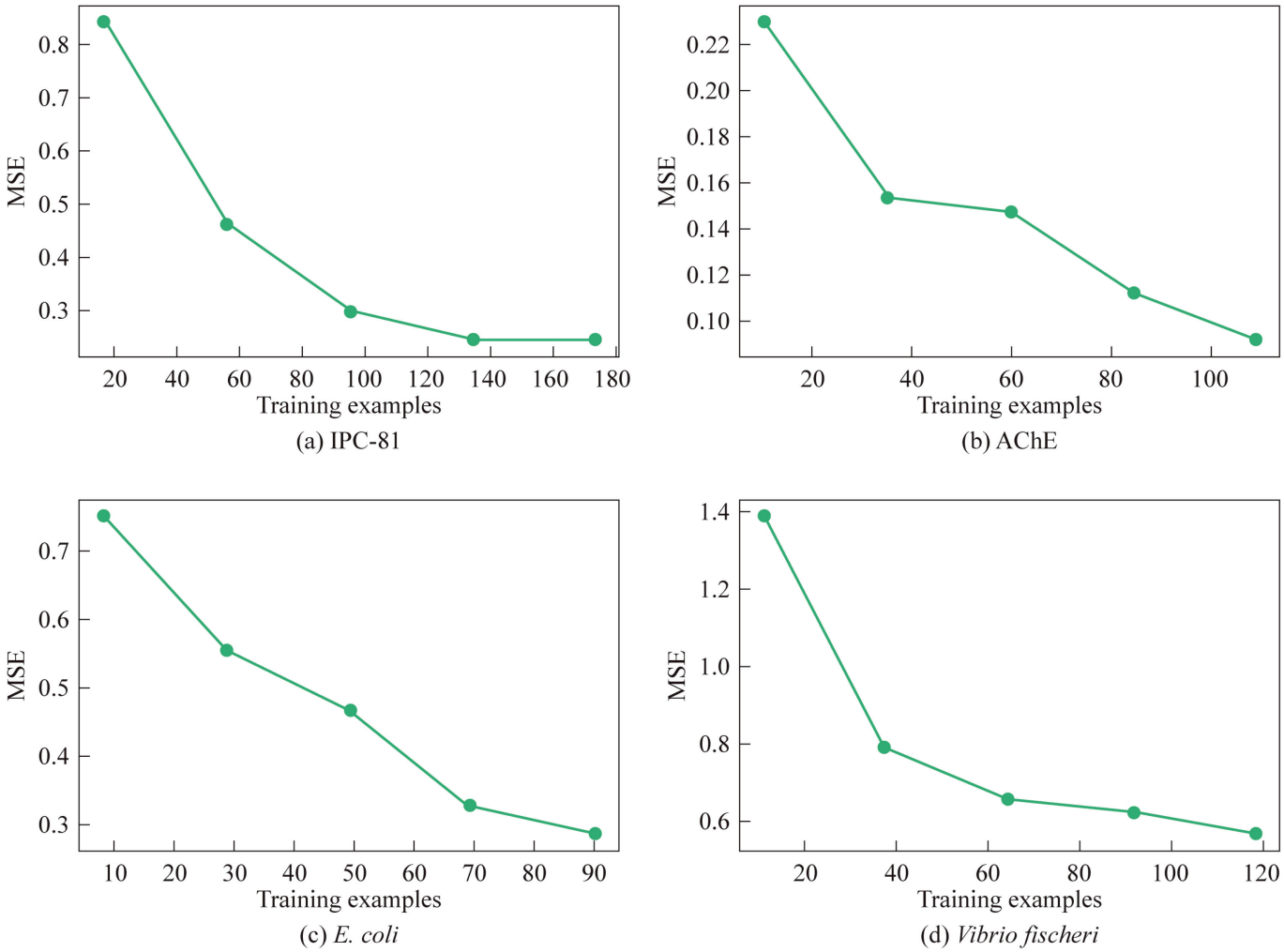

Fig. 3. The learning curves for different models on four datasets with the CatBoost algorithm.

3.4. Model interpretation

A SHAP [29] analysis was utilized to explain the CatBoost model on four different datasets; the importance of various features on ILs' toxicity is illustrated in Fig. 5, the *Y*-axis is the feature name; whereas the *X*-axis is the mean SHAP value averaged over the whole dataset. The larger the mean SHAP value of a feature, the greater its importance to the toxicity of ILs. For dataset IPC-81, features "VSA EState Descriptor 7 (VSA_EState7)", "EState VSA Descriptor 8 (EState_VSA8)", and "MOE MR VSA Descriptor 5 (SMR_VSA5)" play the top important roles on the toxicity. For dataset AChE, the top important features are "MOE MR VSA Descriptor 5 (SMR_VSA5)", "MOE Charge VSA Descriptor 8 (PEOE_VSA8)", and "MOE Charge VSA Descriptor 8 (PEOE_VSA9)". For dataset *E. coli*, "Hall-Kier

Table 1
Metrics of various models in predicting the toxicity of ILs.

Datasets	Metrics	Bagging	AdaBoost	GBoost	Stacking	Voting	CatBoost
IPC-81	MSE	0.259	0.251	0.276	0.319	0.238	**0.230**
	RMSE	0.507	0.498	0.522	0.562	0.486	**0.479**
	MAE	0.390	0.354	0.383	0.409	0.363	**0.357**
	R^2	0.824	0.827	0.809	0.782	0.836	**0.845**
AChE	MSE	0.062	0.050	0.064	0.077	0.051	**0.046**
	RMSE	0.249	0.223	0.251	0.276	0.226	**0.214**
	MAE	0.187	0.169	0.186	0.206	0.173	**0.172**
	R^2	0.756	0.803	0.749	0.697	0.799	**0.819**
E.coli	MSE	0.245	0.238	0.298	0.348	0.237	**0.227**
	RMSE	0.484	0.478	0.528	0.576	0.474	**0.464**
	MAE	0.378	0.350	0.392	0.458	0.363	**0.355**
	R^2	0.849	0.852	0.816	0.782	0.854	**0.859**
Vibrio fischeri	MSE	0.524	0.488	0.580	0.539	0.495	**0.440**
	RMSE	0.715	0.679	0.754	0.729	0.693	**0.655**
	MAE	0.522	0.473	0.541	0.540	0.501	**0.470**
	R^2	0.729	0.758	0.694	0.715	0.747	**0.774**

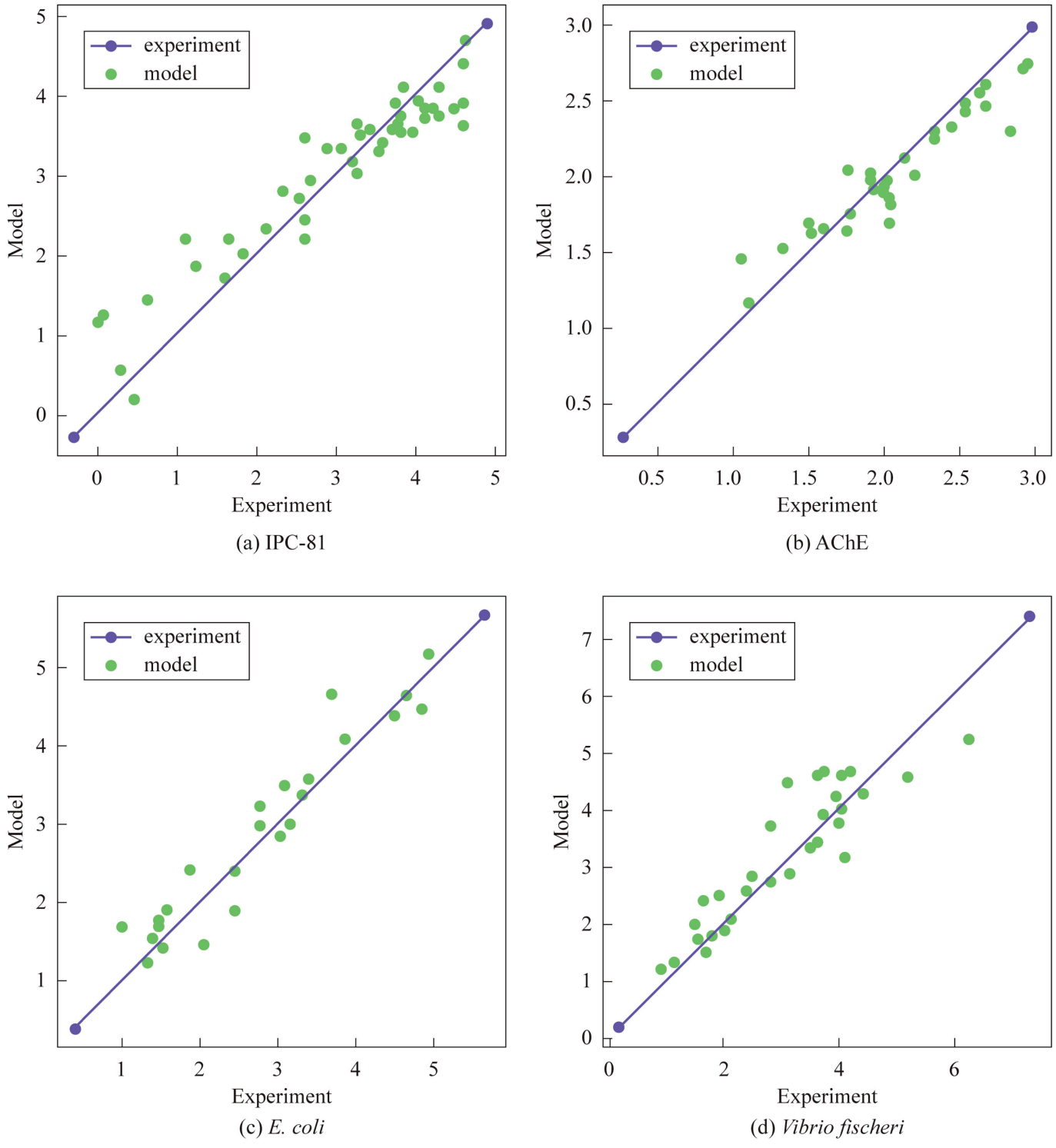

Fig. 4. Parity plot of predicted and experimental values across four datasets (test sets) by the CatBoost model.

Kappa2 value (Kappa2)", "Morgan fingerprint, radius 1 (FpDensity-Morgan1)", and "MOE Charge VSA Descriptor 6 (PEOE_VSA6)" are the top 3 important features. For the dataset *Vibrio fischeri*, "MOE MR VSA Descriptor 5(SMR_VSA5)", "MOE Charge VSA Descriptor 6 (PEOE_VSA6)", and "EState VSA Descriptor 1 (EState_VSA1)" occupy the top three positions of importance.

Top 10 important features with specific mean SHAP value (MSV) are exhibited in Table 3. The features VSA_EState7, SMR_VSA5, and PEOE_VSA6 all appear in three datasets simultaneously, indicating that ILs' toxicity is greatly influenced by these three features. Features EState_VSA8, FpDensityMorgan1, PEOE_VSA8, Kappa2, and EState_VSA1 appear in two datasets, and their importance cannot be ignored.

Merely understanding the importance of features is insufficient; it is also necessary to analyze the positive and negative contributions of these features to ILs' toxicity, which is illustrated in Fig. 6.

Table 2
Performance comparison of various models from previous literatures.

Datasets	Metrics	SVM [43,44]	RF [36,37,44]	MLP [44,45]	XGBoost [27,46]	ANN [27,47]	CatBoost (this work)
IPC-81	MSE	–	–	–	–	–	**0.230**
	RMSE	0.687	0.655	0.640	0.530	0.620	**0.479**
	MAE	0.510	0.436	0.467	–	–	**0.357**
	R^2	0.683	0.706	0.688	0.790	0.710	**0.845**
AChE	MSE	–	–	–	–	–	**0.046**
	RMSE	0.229	0.286	0.243	0.200	0.260	**0.214**
	MAE	0.128	0.170	0.156	–	–	**0.172**
	R^2	0.814	0.746	0.825	0.840	0.720	**0.819**
E. coli	MSE	–	–	–	–	–	**0.227**
	RMSE	0.545	0.547	0.665	0.600	0.670	**0.464**
	MAE	0.396	0.407	0.462	–	–	**0.355**
	R^2	0.832	0.811	0.775	0.810	0.760	**0.859**
Vibrio fischeri	MSE	–	–	–	–	–	**0.440**
	RMSE	0.739	0.724	0.658	0.800	0.800	**0.655**
	MAE	0.524	0.460	0.476	–	–	**0.470**
	R^2	0.738	0.753	0.779	0.720	0.700	**0.774**

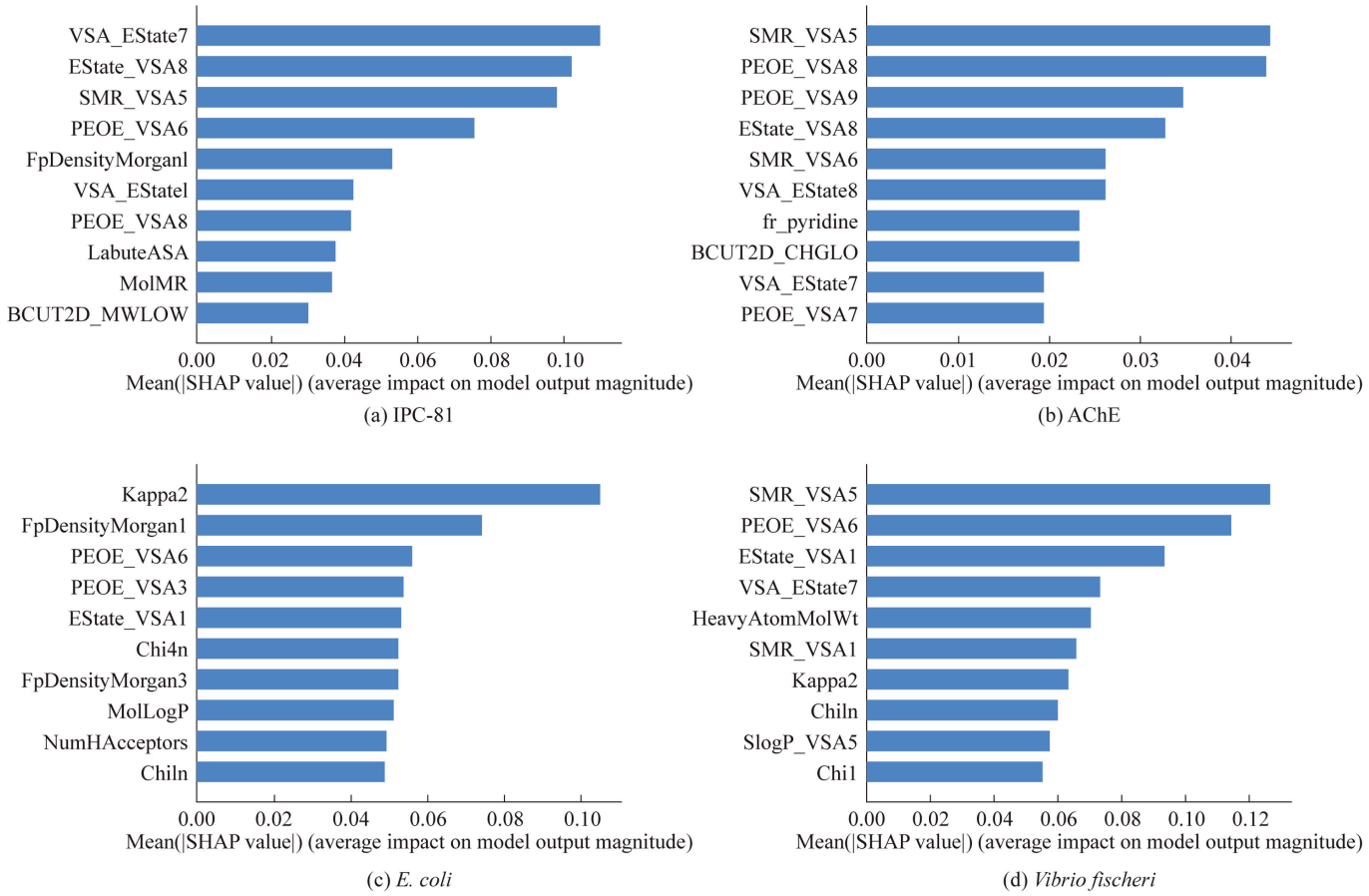

Fig. 5. Feature importance for the CatBoost model on four different datasets.

The *y*-axis is also the feature names in ascending order of significance; while the *x*-axis is the SHAP value. A SHAP value for each feature of an instance is represented by each point on the plot. The color indicates the feature's value, which is overlapped with SHAP values on the *y*-axis; high feature values are shown by the red color while low feature values are shown by the blue color. A SHAP value greater than 0 for each feature indicates that the model utilizes that feature to increase $lgEC_{50}$/lgMIC prediction, which means the reduction of ILs' toxicity; whereas a value less than 0 indicates the enhancement of ILs' toxicity by the feature.

For dataset IPC-81, as shown in Fig. 6(a), the three most important features are VSA_EState7, EState_VSA8, and SMR_VSA5. The larger the value of the feature VSA_EState7, the greater its SHAP value, the lower the toxicity, which indicates that increasing this feature can reduce the toxicity. A high EState_VSA8 feature value might lessen the toxicity as well, while a high SMR_VSA5 feature value will raise the toxicity.

For dataset AChE, as shown in Fig. 6(b), for both features SMR_VSA5 and PEOE_VSA8, increasing their values will enhance the ILs' toxicity. While as the PEOE_VSA9 value rises, the toxicity

Table 3
Mean SHAP value of features on four different datasets.

IPC-81		AChE		*E.coli*		*Vibrio fischeri*	
Feature	MSV	Feature	MSV	Feature	MSV	Feature	MSV
VSA_EState7	0.110	SMR_VSA5	0.045	Kappa2	0.105	SMR_VSA5	0.127
EState_VSA8	0.102	PEOE_VSA8	0.044	FpDensityMorgan1	0.074	PEOE_VSA6	0.114
SMR_VSA5	0.098	PEOE_VSA9	0.035	PEOE_VSA6	0.056	EState_VSA1	0.094
PEOE_VSA6	0.075	EState_VSA8	0.033	PEOE_VSA3	0.054	VSA_EState7	0.074
FpDensityMorgan1	0.053	SMR_VSA6	0.026	EState_VSA1	0.053	HeavyAtomMolWt	0.071
VSA_EState1	0.043	VSA_EState8	0.026	Chi4n	0.053	SMR_VSA1	0.067
PEOE_VSA8	0.042	fr_pyridine	0.024	FpDensityMorgan3	0.053	Kappa2	0.064
LabuteASA	0.038	BCUT2D_CHGLO	0.024	MolLogP	0.051	Chi1n	0.060
MolMR	0.037	VSA_EState7	0.020	NumHAcceptors	0.050	SlogP_VSA5	0.058
BCUT2D_MWLOW	0.030	PEOE_VSA7	0.020	Chi1n	0.049	Chi1	0.056

value of ILs falls. The feature “Number of pyridine rings (fr_pyridine)" has a positive impact on toxicity, indicating that an IL molecule's toxicity will rise as the number of pyridine rings increases. Because of its extreme hydrophobicity, the pyridine ring can break down cell membrane structure, making ILs more hazardous. This is consistent with the experimental findings [5–7].

For dataset *E. coli*, features that have positive contributions to the toxicity are Kappa2 and PEOE_VSA6, while feature FpDensityMorgan1 contributes negatively to toxicity. The NumHAcceptors feature “The number of Hydrogen Bond Acceptors” also contributes negatively to toxicity, which means that the more Hydrogen Bond Acceptors an IL molecule has, the less toxicity it has. Hydrogen Bond Acceptors can form hydrogen bonds to enhance hydrophilicity, which will effectively reduce the ILs' toxicity.

For dataset *Vibrio fischeri*, all the three most important features SMR_VSA5, PEOE_VSA6, and EState_VSA1 have positive contributions to the toxicity. Feature “The average molecular weight of the molecule ignoring hydrogens (HeavyAtomMolWt)" also has a positive effect. As the average molecular weight of ILs rises, so does their toxicity.

We also found that on different datasets, the trend of the same feature's contribution to toxicity is the same. VSA_EState7 has a

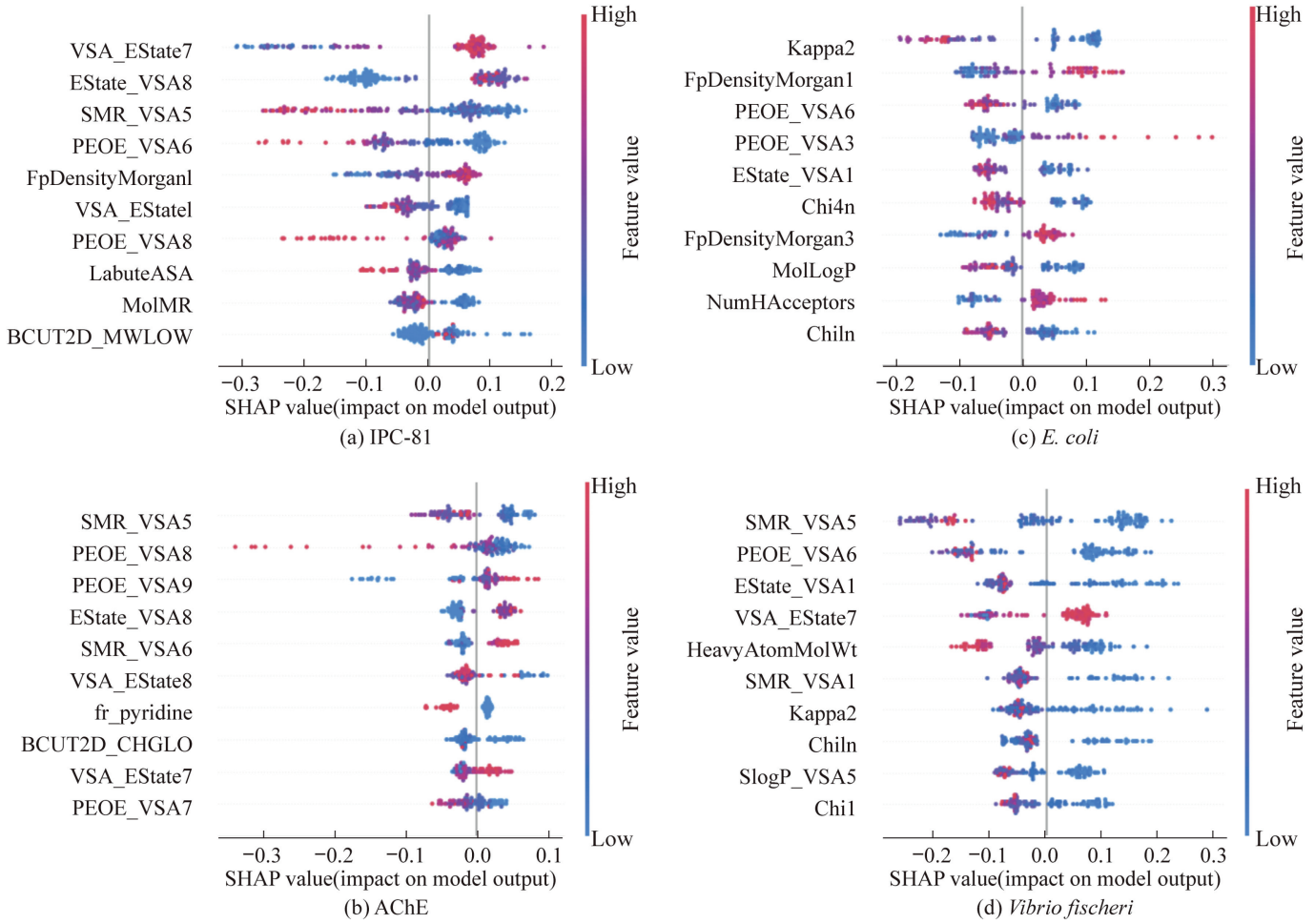

Fig. 6. Positive and negative contributions of features for the CatBoost model on four different datasets.

negative contribution to toxicity on datasets IPC-81, AChE and *Vibrio fischeri*. SMR_VSA5 contributes positively to toxicity on datasets IPC-81, AChE and *Vibrio fischeri*. The toxicity of the IPC-81, *E.coli* and *Vibrio fischeri* datasets is positively impacted by PEOE_VSA6.

The important finding of this work is that, through a data-driven approach, we have discovered the positive and negative impacts of descriptors on ILs' toxicity. Using the CatBoost [34,35] model, we provide a microscopic explanation of ILs' toxicity, offering a theoretical basis for the rapid screening and development of more environmentally-benign and low-toxicity ILs. However, although this work used four different datasets, each dataset is still quite small, and the model's accuracy and generalization ability need further improvement. Additionally, the factors influencing the toxicity of ILs are quite complex. This work only considered molecular descriptors and did not take into account the interactions between cations and anions that might affect toxicity, which is an area that needs improvement in future works.

4. Conclusions

Six different machine learning models (Bagging, AdaBoost, GBoost, Stacking, Voting and CatBoost) are built to predict the toxicity of ILs on four datasets (IPC-81, AChE, *E. coli*, *Vibrio fischeri*). Molecular descriptors produced from SMILES structure serve as the input features. All models are evaluated using MSE, RMSE, MAE and R^2 to achieve the best model. Furthermore, an interpretation model based on SHAP analysis is established to discover how each molecular feature affects toxicity, both positively and negatively. With the lowest MSE, RMSE, MAE and the highest R^2, the CatBoost model performs best among all models. The CatBoost model has additional parameters and complexity, which make it a more reliable model in predicting the toxicity of ILs. The model's interpretation results show that SMR_VSA5, SMR_VSA5, PEOE_VSA8, Kappa2, PEOE_VSA6, SMR_VSA5, PEOE_VSA6 and EState_VSA1 are the most important positive features, which can raise the toxicity of IL as their levels rise; while VSA_EState7, EState_VSA8, PEOE_VSA9 and FpDensityMorgan1 are the most important negative features, which can reduce the toxicity as their levels increase. Plus, the rising of both the number of pyridine rings and the average molecular weight of an IL will increase its toxicity, whereas the increasing of Hydrogen Bond Acceptors in an IL molecule will lessen its toxicity. Our findings offer new insights on IL design; the accurate and interpretable model can be utilized to rapidly screen IL with low toxicity for industrial applications.

CRediT Authorship Contribution Statement

Haijun Feng: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Data curation. Li Jiajia: Writing – review & editing, Validation. Zhou Jian: Writing – review & editing, Supervision, Investigation.

Declaration of Competing Interest

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

Acknowledgements

This study was funded by Research Platforms and Projects for Higher Education Institutions of Department of Education of Guangdong Province in 2024 (2024KTSCX256), and 2023 Guangdong Province Higher Vocational Education Teaching Quality and Teaching Reform Project (2023JG080).

References

[1] M.Y. Zhu, Y. Yang, Poly(ionic liquid)s: an emerging platform for green chemistry, *Green Chem.* 26 (9) (2024) 5022–5102.
[2] L. Zhang, J. Liu, S. Lou, H. Jiang, S. Wang, S. Li, T. Tang, A. Zhang, Synergistic flame-retardant epoxy resin using quaternary phosphate naphthalene sulfonate ionic liquid and dimethyl methylphosphonate, *Chem. J. Chin. Univ.* 45 (4) (2024) 17–26. (in Chinese)
[3] F. Li, T. Zhang, L. Lv, W.X. Tang, Y. Wang, S.W. Tang, Effects of ionic liquids on the vapor–liquid equilibriumof 1, 3, 5-trioxane–water system at 101.3 kPa, *Chin. J. Chem. Eng.* 73 (2024) 42–50.
[4] P.M. Probert, A.C. Leitch, M.P. Dunn, S.K. Meyer, J.M. Palmer, T.M. Abdelghany, A.F. Lakey, M.P. Cooke, H. Talbot, C. Wills, W. McFarlane, L.I. Blake, A.K. Rosenmai, A. Oskarsson, R. Figueiredo, C. Wilson, G.E. Kass, D.E. Jones, P.G. Blain, M.C. Wright, Identification of a xenobiotic as a potential environmental trigger in primary biliary cholangitis, *J. Hepatol.* 69 (5) (2018) 1123–1135.
[5] A. Jakubowska, E. Grabińska-Sota, Study of the toxicity of five quaternary ammonium ionic liquids to aquatic organisms, *Desalination Water Treat.* 117 (2018) 202–210.
[6] G. Jeremias, F. Jesus, S.P.M. Ventura, F.J.M. Gonçalves, J. Asselman, J.L. Pereira, New insights on the effects of ionic liquid structural changes at the gene expression level: molecular mechanisms of toxicity in Daphnia Magna, *J. Hazard. Mater.* 409 (2021) 124517.
[7] C. Zhang, L.S. Zhu, J.H. Wang, J. Wang, T.T. Zhou, Y.Q. Xu, C. Cheng, The acute toxic effects of imidazolium-based ionic liquids with different alkyl-chain lengths and anions on zebrafish (Danio rerio), *Ecotoxicol. Environ. Saf.* 140 (2017) 235–240.
[8] D.M. Makarov, Y.A. Fadeeva, E.A. Safonova, L.E. Shmukler, Predictive modeling of antibacterial activity of ionic liquids by machine learning methods, *Comput. Biol. Chem.* 101 (2022) 107775.
[9] M. Villanueva, P. Vallet, T. Teijeira, A. Santiago-Alonso, A. Amigo, E. Tojo, L.M. Varela, J.J. Parajó, J. Salgado, Effect of alkyl chain length on the thermal properties and toxicity of n-alkyl-ammonium nitrate ionic liquids ($n = 2, 3, 4, 5, 6, 8$) for energy applications, *J. Therm. Anal. Calorim.* 150 (2025) 6851–6861.
[10] H.J. Feng, L.L. Qin, B.X. Zhang, J. Zhou, Prediction and interpretability of melting points of ionic liquids using graph neural networks, *ACS Omega* 9 (14) (2024) 16016–16025.
[11] V. Venkatraman, S. Evjen, H.K. Knuutila, A. Fiksdahl, B.K. Alsberg, Predicting ionic liquid melting points using machine learning, *J. Mol. Liq.* 264 (2018) 318–326.
[12] H.J. Feng, W. Qin, G.W. Hu, H.J. Wang, Intelligent prediction of nitrous oxide capture in designable ionic liquids, *Appl. Sci.* 13 (12) (2023) 6900.
[13] H.J. Feng, P.G. Zhang, W. Qin, W.M. Wang, H.J. Wang, Estimation of solubility of acid gases in ionic liquids using different machine learning methods, *J. Mol. Liq.* 349 (2022) 118413.
[14] Y. Jian, Y.Y. Wang, A. Barati Farimani, Predicting CO_2 absorption in ionic liquids with molecular descriptors and explainable graph neural networks, *ACS Sustainable Chem. Eng.* 10 (50) (2022) 16681–16691.
[15] D.M. Makarov, Y.A. Fadeeva, V.A. Golubev, A.M. Kolker, Designing deep eutectic solvents for efficient CO_2 capture: a data-driven screening approach, *Sep. Purif. Technol.* 325 (2023) 124614.
[16] F.Z. Meng, S.Y. Han, L. Lin, J.L. Li, K.L. Chen, J.G. Jiang, Process optimization and mechanism study of ionic liquid-based mixed amine biphasic solvents for CO_2 capture in biogas upgrading procedure, *Front. Environ. Sci. Eng.* 18 (8) (2024) 95.
[17] S.J. Zeng, X.Q. Sun, Y.G. Bai, L. Bai, S. Zheng, X.P. Zhang, S.J. Zhang, Research progress of CO_2 capture and separation by functionalized ionic liquids and materials, *Acta Chim. Sin.* 81 (6) (2023) 627.
[18] Y.X. Fu, X.Y. Liu, J.Z. Gao, Y. Lei, Y.Q. Chen, X.P. Zhang, Machine learning models for the density and heat capacity of ionic liquid–water binary mixtures, *Chin. J. Chem. Eng.* 73 (2024) 244–255.
[19] Y. Lei, Y. Shu, X.G. Liu, X.Y. Liu, X.Q. Wu, Y.Q. Chen, Predictive modeling on the surface tension and viscosity of ionic liquid-organic solvent mixtures *via* machine learning, *J. Taiwan Inst. Chem. Eng.* 151 (2023) 105140.
[20] Y.Q. Chen, S.L. Ma, Y. Lei, X.D. Liang, X.Y. Liu, G.M. Kontogeorgis, R. Gani, Ionic liquid binary mixtures: machine learning-assisted modeling, solvent tailoring, process design, and optimization, *AIChE J.* 70 (5) (2024) e18392.
[21] N.H. Abdullah, D. Zaini, B. Lal, Prediction of ionic liquids toxicity using machine learning models for application to gas hydrate, *Process, Saf. Prog.* 43 (2024) S199–S212.
[22] D.C. Fan, K. Xue, R.Q. Zhang, W.G. Zhu, H.R. Zhang, J.G. Qi, Z.Y. Zhu, Y.L. Wang, P.Z. Cui, Application of interpretable machine learning models to improve the prediction performance of ionic liquids toxicity, *Sci. Total Environ.* 908 (2024) 168168.
[23] Y.S. Zhao, J.H. Zhao, Y. Huang, Q. Zhou, X.P. Zhang, S.J. Zhang, Toxicity of ionic liquids: database and prediction *via* quantitative structure-activity relationship method, *J. Hazard. Mater.* 278 (2014) 320–329.
[24] A. Sosnowska, M. Grzonkowska, T. Puzyn, Global versus local QSAR models for predicting ionic liquids toxicity against IPC-81 leukemia rat cell line: the predictive ability, *J. Mol. Liq.* 231 (2017) 333–340.

[25] L.D. Cao, P. Zhu, Y.S. Zhao, J.H. Zhao, Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids, *J. Hazard. Mater.* 352 (2018) 17–26.
[26] S.Y. Ma, M. Lv, X.Y. Zhang, H.L. Zhai, W.J. Lv, Computational study of the effects of cations and anions to the cytotoxicity of diverse ionic liquids by supervised machine learning, *Chemometr. Intell. Lab. Syst.* 144 (2015) 138–147.
[27] J.C. Yan, G.H. Liu, H.L. Chen, S. Hu, X.H. Wang, B. Yan, X.L. Yan, ILTox: a curated toxicity database for machine learning and design of environmentally friendly ionic liquids, *Environ. Sci. Technol. Lett.* 10 (11) (2023) 983–988.
[28] J.C. Yan, X.L. Yan, S. Hu, H. Zhu, B. Yan, Comprehensive interrogation on acetylcholinesterase inhibition by ionic liquids using machine learning and molecular modeling, *Environ. Sci. Technol.* 55 (21) (2021) 14720–14731.
[29] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using shapley additive explanations (SHAP), *Case Stud. Constr. Mater.* 16 (2022) e01059.
[30] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (2) (1996) 123–140.
[31] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (1) (1997) 119–139.
[32] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Stat.* 29 (5) (2001) 1189–1232.
[33] D.H. Wolpert, Stacked generalization, *Neural Netw.* 5 (2) (1992) 241–259.
[34] A.V. Dorogush, V. Ershov, A. Gulin, Catboost: Gradient Boosting with Categorical Features Support, 2018, https://doi.org/10.48550/arXiv.1810.11363.
[35] J.T. Hancock, T.M. Khoshgoftaar, CatBoost for big data: an interdisciplinary review, *J. Big Data* 7 (1) (2020) 94.
[36] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach. Learn* 63 (1) (2006) 3–42.
[37] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
[38] P. Labute, A widely applicable set of descriptors, *J. Mol. Graph. Model.* 18 (4–5) (2000) 464–477.
[39] B. Ramsundar, P. Eastman, P. Walters, V. Pande, Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More, O'Reilly Media Inc., 2019.
[40] Molecular descriptors, https://datagrok.ai/help/datagrok/solutions/domains/chem/descriptors.
[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
[42] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, G. Varoquaux, API Design for Machine Learning Software: Experiences from the scikit-learn Project, 2013. https://doi.org/10.48550/arXiv.1309.0238.
[43] R. Fan, K. Chang, C. Hsieh, X. Wang, C. Lin, LIBLINEAR: a library for large linear classification, *Mach. Learn. Res.* 9 (2008) 1871–1874.
[44] H. Feng, B. Zhang, J. Zhou, Predicting and interpreting the toxicity of ionic liquids using graph neural network", *CIESC Journal* 76 (2025) 93–106. (in Chinese)
[45] G.E. Hinton, Connectionist learning procedures, *Artif. Intell.* 40 (1–3) (1989) 185–234.
[46] T.Q. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, 2016.
[47] A. Krogh, What are artificial neural networks? *Nat. Biotechnol.* 26 (2008) 195–197.