**1. The exact hyperparameters of the used ML methods**

For the Random Forest (RF) and XGBoost, the grid search method was applied to optimize a best parameter set for each model. Specifically, for the RF, only one parameter, mtry was optimized; for the XGBoost, four important parameters, Eta, Gamma, max\_depth and min\_weight were optimized. The optimization results were as follows:

RF: mtry=100;

XGBoost: Boosting rounds=2000;

Eta=0.3;

Gamma=0;

max\_depth=6;

min\_weight=1.

**2. The data splits and model selection method**

In this manuscript, we chose the best models based on their predictive power. To evaluate the predictive performance of obtained models, the Monte-Carlo cross validation was employed to evaluate the model performance. For each evaluation process, the data splits method was as follows: 80% compounds were randomly chosen to build models and the remaining 20% were used as the test set. This process was repeated 100 times and their average accuracy values were taken as the assessment indexes to select the best ones for final application.

Five common statistical parameters were used to evaluate the performances of QSAR models: sensitivity (SE), specificity (SP), accuracy (ACC), F value (F), an area under receiver operating characteristic curve (AUC).