

Supporting information

The SMILES string, classification, CAS number and names for the training set and external validation set were listed in Table SI1. The parameters ranges in grid search for different machine learning method were shown in Table SI2. The names for the 49 descriptors were listed in Table SI3. The performance of the models for five-fold cross validation and external validation using different algorithms combined with different fingerprints or descriptors was listed in Table SI4. The correlation heat map of 49 descriptors was shown in Figure SI1.