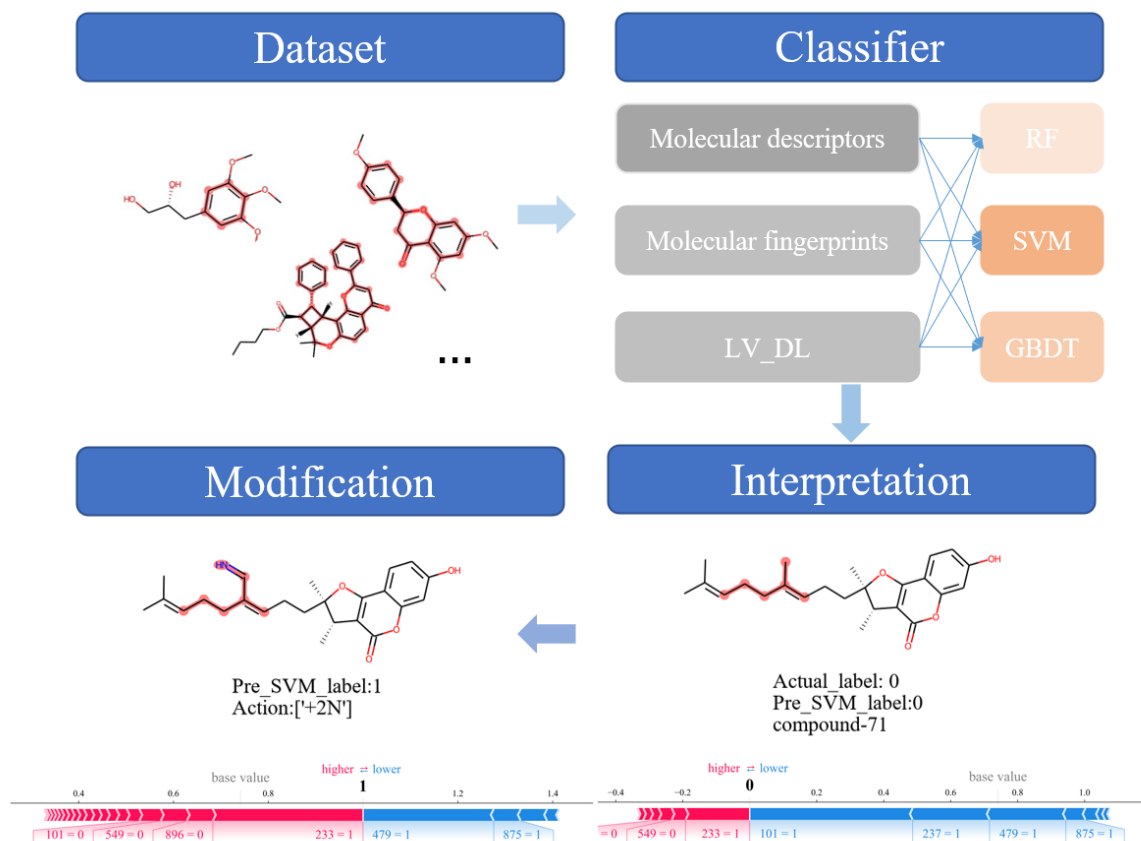


Prediction and interpretation microglia cytotoxicity by machine learning

This is the code for "Prediction and interpretation microglia cytotoxicity by machine learning".



Requirements

- Python 3.6.10
- [rdkit 2021.03.5](#)
- [pytorch 1.9.0](#)
- Other dependencies noted in `requirements.txt`, please install them with `pip`.

Demo

Step1. Data preprocess. The molecular fingerprints, molecular descriptors and LV_DL under random and murcko generic scaffold data split strategies were obtained respectively.

```
python dataset_preprocess_random.py
python dataset_preprocess_BM.py
```

Step2. Construct cytotoxicity model. The molecular fingerprints, molecular descriptors and LV_DL coupled with RF, SVM and GBDT modeling methods to construct cytotoxicity model. The predictive accuracy (Acc), balanced accuracy (BA) and F1-score were applied for evaluating the model performance.

```
python APM_model_random.py
python APM_model_BM.py
```

Step3. Feature selection. RFE-SVC feature selection method were used to Avalon, ECFP4, FCFP4, RDkit molecular fingerprints. And the molecular fingerprint after feature selection and three modeling methods are used to construct the model and analyze.

```
python FS_overall.py
python FS_RFE_random.py
python APM_model_FS_random.py
python APM_model_BM_FS.py
```

Step4. SHAP analysis. Global model performance and SAs analysis.

```
python shap_interpreter_three_model.py
python plot_inter5_substructure.py
python shap_typical_plot.py
```

Step5. Molecular cytotoxic modification strategy . The important cytotoxic substructures can be obtained by SHAP, which were modified in a chemically valid way by adding atoms or bonds and removing bonds.

```
python RL_single_step.py
python Direct_two_step.py
python read_file.py
python SHAP_analysis_compound-17.py
```

Results

The model performance results, SHAP analysis results and molecular cytotoxic modification results can be gained in file folder "data".