# Steps to generate predictions for a new list of chemicals

## Final Models Parameters

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Fraction Unbound in Plasma** | **Intrinsic Clearance**  **(Classification)** | **Intrinsic Clearance**  **(Regression)** |
| **Fingerprints** | PubChem  (Using KNIME version 3.7.1)  +  ToxPrints  (Using Chemotyper version1.0.r12976) | PubChem  (Using KNIME version 3.7.1)  +  ToxPrints  (Using Chemotyper version1.0.r12976) | PubChem  (Using KNIME version 3.7.1)  +  ToxPrints  (Using Chemotyper version1.0.r12976) |
| **Descriptors** | OPERA 2.5  (LogP and pKa) | OPERA 2.5  (LogP and pKa) | - |
| **Algorithm** | Consensus of SVM and Random Forest | SVM | Random Forest |
| **Final Features to the Model** | Human.Funbound.plasma\_Features.csv | Clint\_Features\_Classification.csv | Clint\_Features\_Regression.csv |
| **Final Model (pickle file)** | fub\_svr.sav fub\_rf.sav | clintClas\_svc.sav | clintReg\_rf.sav |

## 1. Generate sdf file from DSSTox

-Run the script GetQSARreadySDFfileFromDSSTox.py

-Input file: List of CASRNs for which the predictions need to be made

-Output file: QSAR-ready structure (sdf) file

\*This script can be run by an internal EPA user only. Please contact judson.richard@epa .gov or [prachipradeep.15@gmail.com](mailto:prachipradeep.15@gmail.com) for help.

## 2. Generate Pubchem fingerprints and CDK descriptors from KNIME

-Run the KNIME workflow GeneratePubChem&CDK.knwf

-Input file: QSAR-ready structure (sdf) file

-Output file: PubChem fingerprints and CDK descriptors

## 3. Generate OPERA descriptors

-Run the OPERA tool. Select LogP and pKa

-Input file: QSAR-ready structure (sdf) file

-Output file: OPERA descriptors

## 4. Generate predictions

-Script: PradeepFubClintPredictions.py

-Input files: PubChem fingerprints, ToxPrints, OPERA descriptors, pickled model files, final feature sets for each model, OPERA descriptors normalization vector based on training data (opera\_scaler.sav)

-Output file: Fraction Unbound and Intrinsic Clearance Predictions