**User Manual for “SteroidXtract.py”**

(Version 1.0, Sep 30, 2020)

Shipei Xing, Tao Huan\*

Department of Chemistry, Faculty of Science, University of British Columbia, Vancouver Campus, 2036 Main Mall, Vancouver, V6T 1Z1, BC, Canada

\* Author to whom correspondence should be addressed:

Dr. Tao Huan

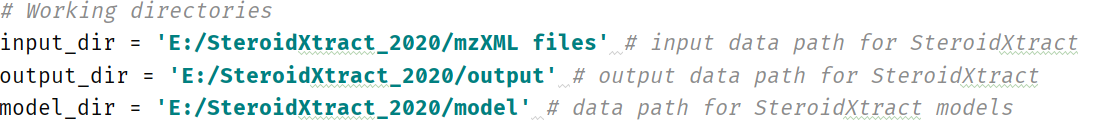
Tel: (+1)-604-822-4891

E-mail: thuan@chem.ubc.ca

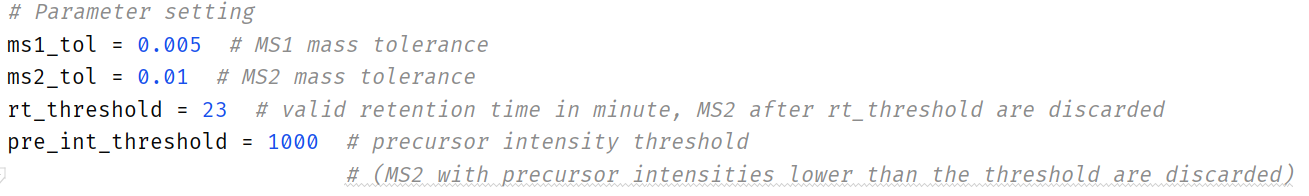
Internet: <https://huan.chem.ubc.ca/>

* This script performs steroid MS2 spectra extraction from untargeted metabolomics datasets.
* The program is written in Python and is publicly available at https://github.com/HuanLab/SteroidXtract.
* Please see below for the code instructions.

1. Install the Python language and following packages: pyteomics, numpy, pandas, os, math, tensorflow.
2. Download and open the Python script “SteroidXtract.py” in a Python IDE (Integrated Development Environment) such as PyCharm or Visual Studio Code.
3. Change the working directories in the Python script. The model directory should contain files “SteroidXtract\_model.json” and “SteroidXtract\_model.h5”. Use “/” instead of “\” in the directory.



1. Set the parameters in the following table.



**Table.** Parameter settings.

|  |  |
| --- | --- |
| **Parameter name** | **Function** |
| ms1\_tol | Mass tolerance for MS1 scans, default 0.005 *m/z*. |
| ms2\_tol | Mass tolerance for MS1 scans, default 0.01 *m/z*. |
| rt\_threshold | Retention time threshold (in minute). MS2 generated after the rt\_threshold will be discarded. |
| pre\_int\_threshold | Precursor intensity threshold. MS2 with precursor intensities lower than pre\_int\_threshold are discarded. |

1. Run the Python script. For each sample, a CSV file containing the prediction results of all the MS2 spectra and a MGF file containing the steroid MS2 spectra will be generated.