**User’s Manual**

1. **Software Background**

Estrogenic effects have been widely observed in the environment, increasing the risk of reproductive disorders and certain cancers. Given the complex cocktails in real-world, effect-based methods (EBM), in conjunction with high resolution mass spectrometry (HRMS)-based nontarget analysis (NTA), present a promising approach for the identification of unknown estrogenic agonists. Previous studies successfully identified unmonitored estrogenic agonists in sewages using EBM-NTA, such as arenobufagin, loratadine and benzophenone. However, a substantial gap still persists between the complex mixtures found in the sewage and estrogenic effects. The key estrogenic agonists and their transformation patterns remain unclear within complex sewage treatment processes, which impedes the risk assessment and precise control of estrogenic effects.

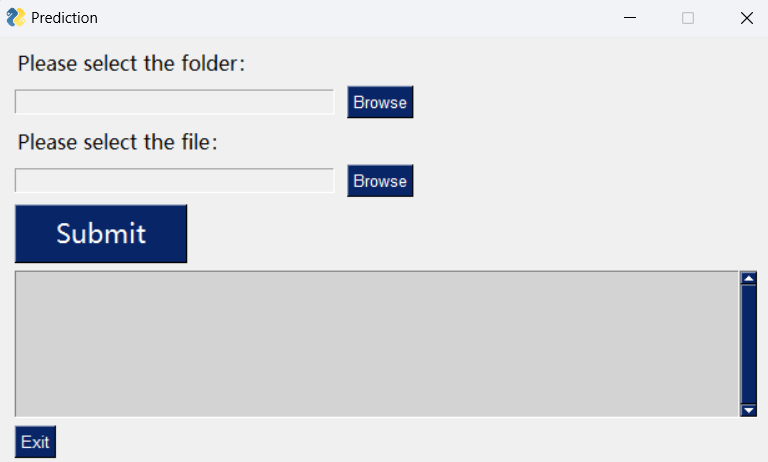
1. **Software Objectives**

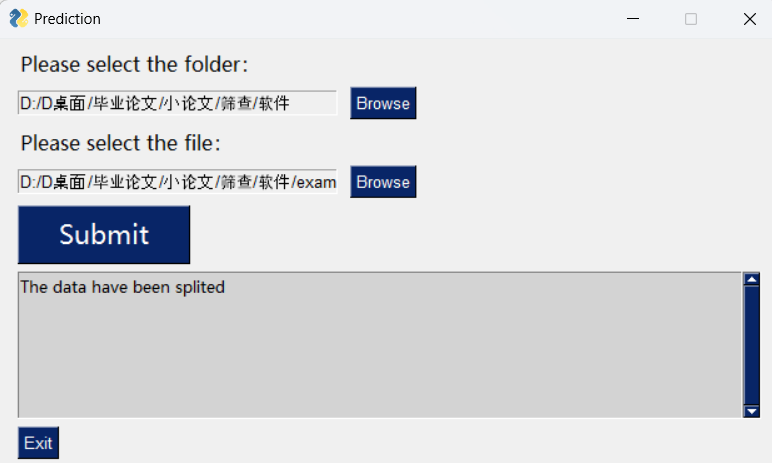
The GNN-FT software is designed to predict the estrogenic activity of compounds based on high-resolution tandem mass spectrometry (HRMS/MS) data. The software integrates a Graph Neural Network (GNN) model with a Fragment-Tree (FT) representation to enhance the identification and prediction of estrogenic agonists in complex mixtures, such as environmental samples. The main objectives of the GNN-FT software are: (1) To bridge HRMS/MS spectral features with chemical bioactivity predictions. (2) To uncover unidentified estrogenic compounds efficiently in real-world samples.

1. **Usage Steps**

The prediction workflow of the GNN-FT model includes the following steps: (1) Converting high-resolution tandem mass spectra of multiple compounds into individual compound files. (2) Generating the fragmentation trees for each compound using Sirius. (3) Compiling the fragmentation trees files and applying the pre-trained GNN model for prediction. (4) Saving the prediction results as an Excel file.

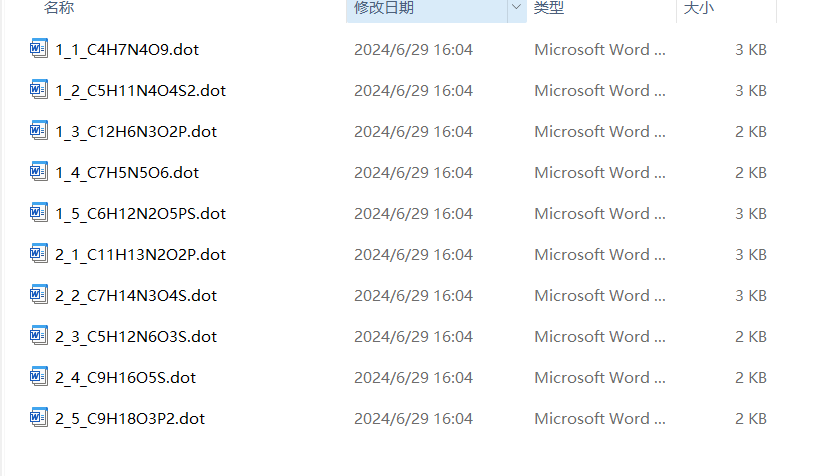
The required input file format for the software is an MGF file containing high-resolution tandem mass spectra. Each compound's tandem spectrum is enclosed between 'BEGIN IONS' and 'END IONS', with the file including information such as the precursor mass (permass), MS level, charge state, and the m/z ratios and intensities of all fragment ions. Save the MGF file of the compound's mass spectrometry secondary fragments that need to be predicted to the folder where the EXE file is located. Click ‘GNN-FT-model.exe’ file. Then click the 'submit' button. After the file conversion is completed, 'The data have been split' will be displayed in the output box. After completion, click the 'Exit' button to exit.



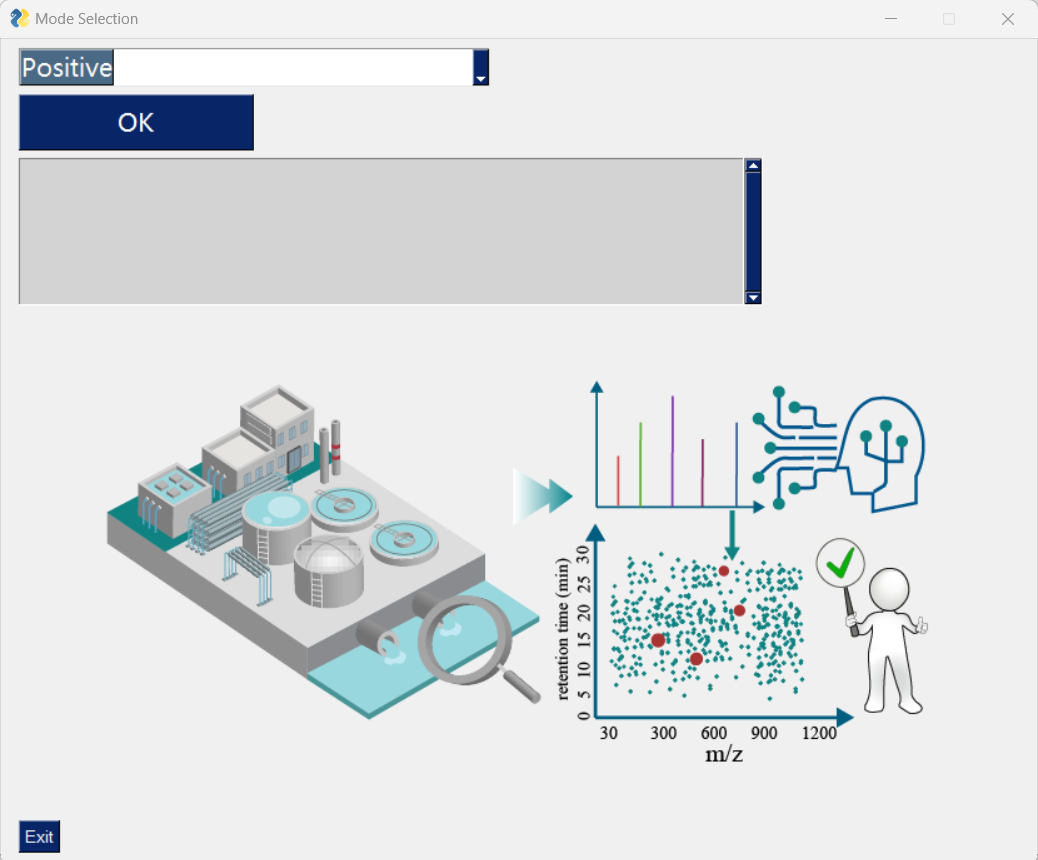


(1) If it is in positive ion mode, please click the run-ff0.bat file in the ./split/FT folder. This will automatically calculate the compound in the MGF file into fragment tree results. The fragment tree results will be in the ./split/FT/tree folder.

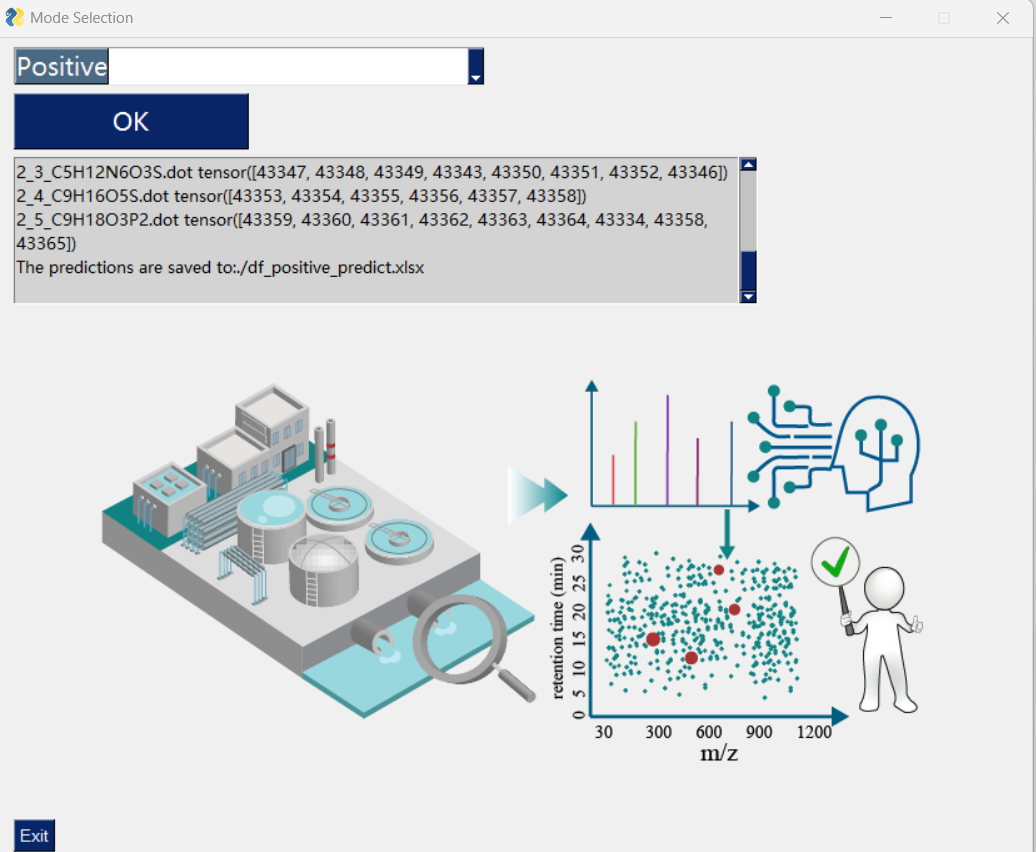




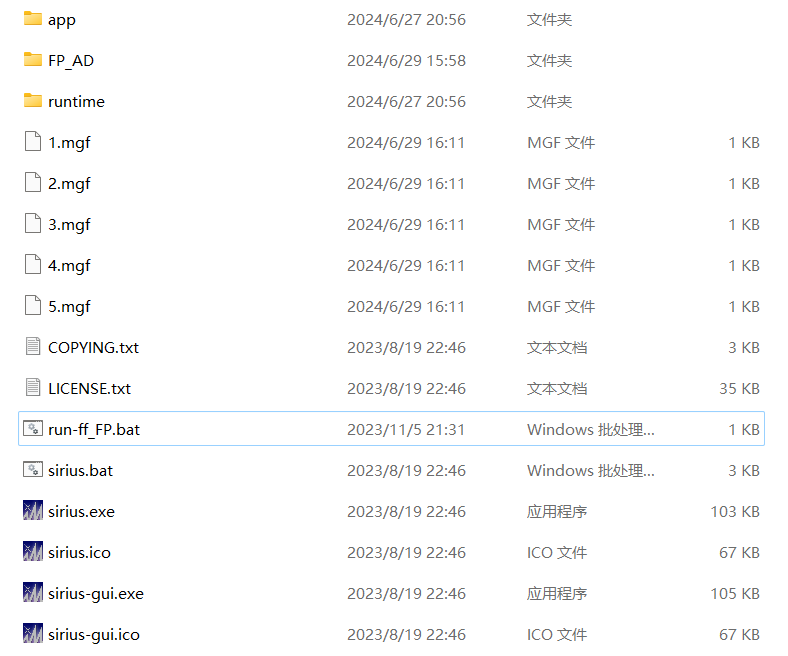
Then, in the pop-up interface, select Positive mode from the dropdown menu and click the 'OK' button to submit to the model for prediction.



After the model prediction is complete, 'The predictions are saved to: ./df\_positive\_predict.xlsx' will appear in the output box.

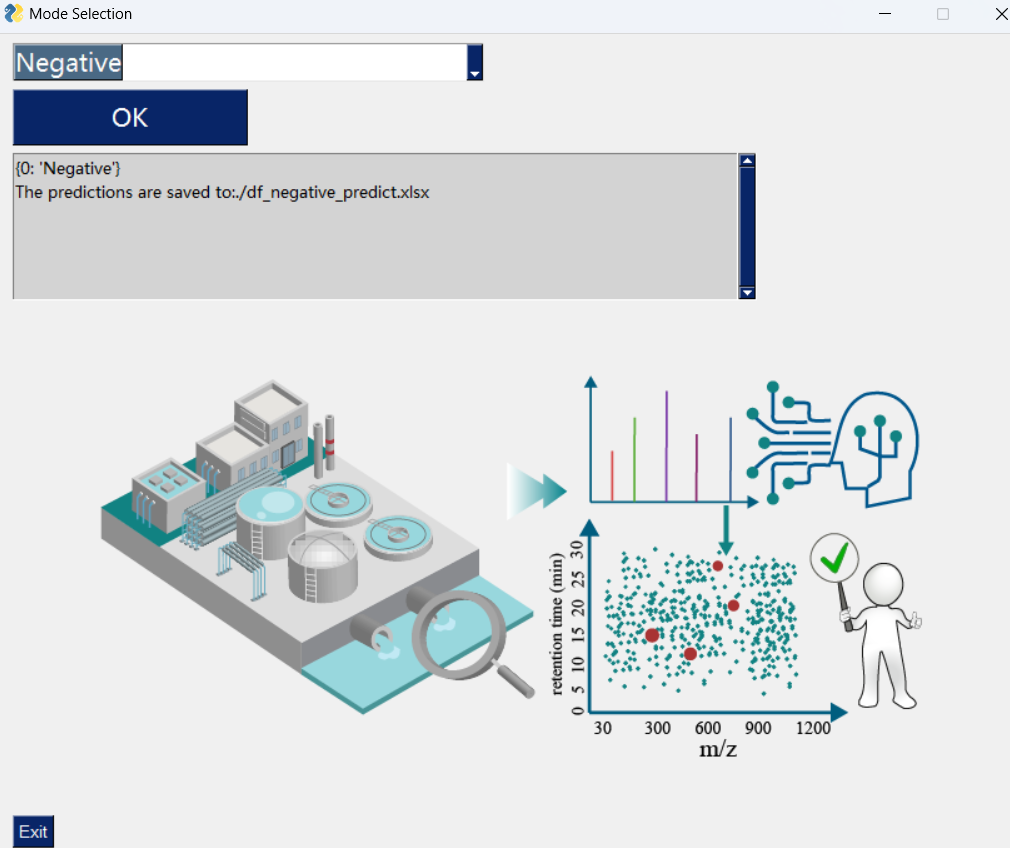


(2) If it is in negative ion mode, please click the run-ff\_FP.bat file in the ./split/FP folder. If a password is required, press the enter key. This will automatically predict the compounds in the MGF file as fingerprints. The fingerprint prediction results will be in the ./split/FP/FP\_AD folder.





Then, in the pop-up interface, select Negative mode from the dropdown menu and click the 'OK' button to submit to the model for prediction. After the model prediction is complete, 'The predictions are saved to: ./df\_negative\_predict.xlsx' will appear in the output box.



Note: Before starting a new prediction task, make sure to delete the generated MGF file, fragment tree files, and fingerprint prediction result files.