PanCan Spectrum

User's Manual

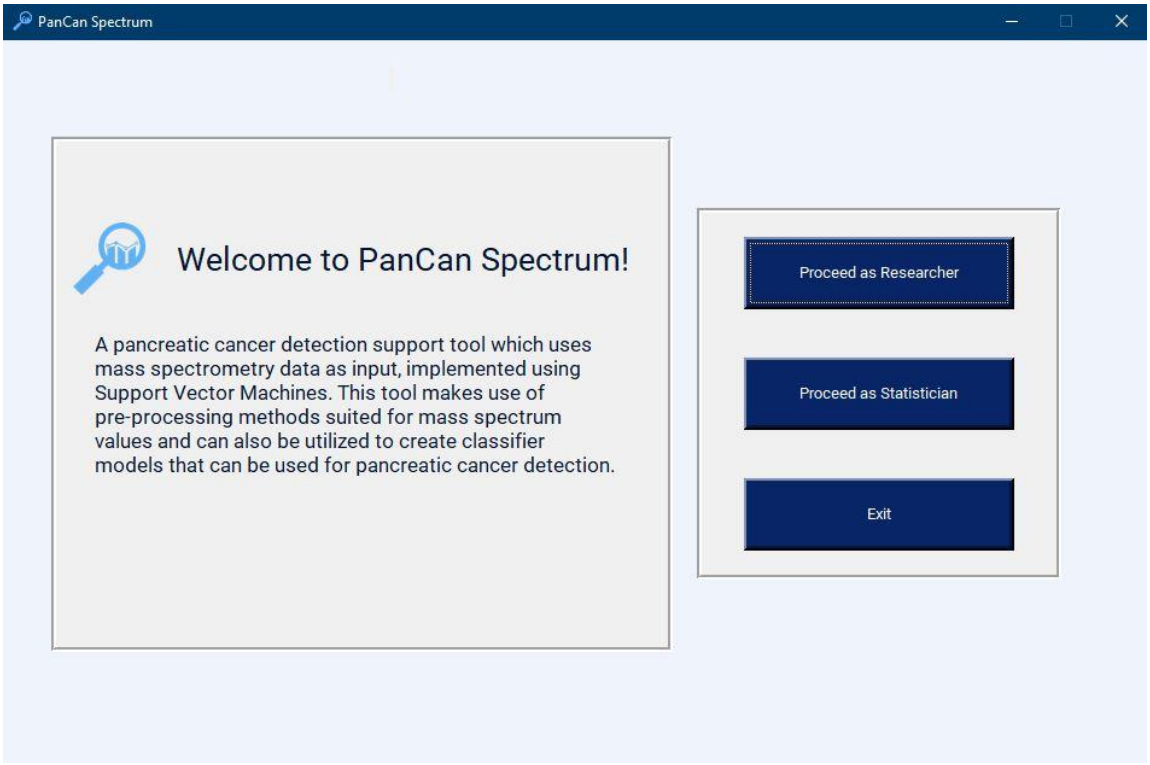
Submitted by: Emmanuel A. Briones

User's Manual

Table of Contents

I. General Information & Setup

A. System Overview



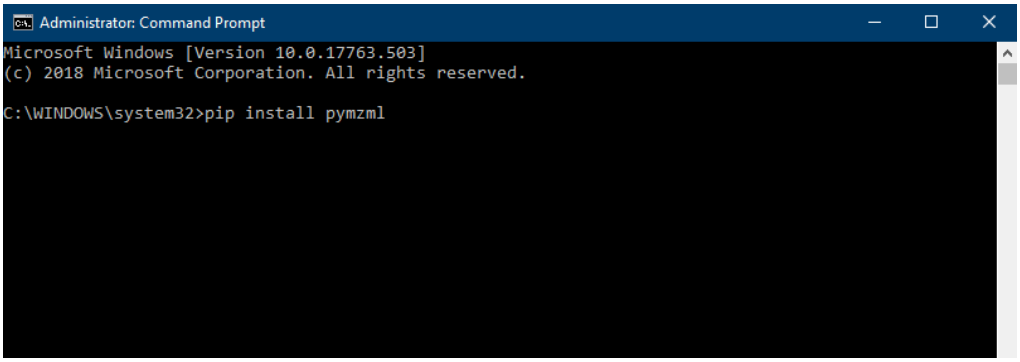
PanCan Spectrum is a simple mass spectrometry-focused tool which utilizes Support Vector Machines algorithm in order to classify MS data based on mass-per-charge values, its corresponding intensities, and retention time. It allows the user to perform varying types of pre-processing methods which are recommended specifically for MS data. Furthermore, it provides the visualization of a mass spectrum, complemented by its data in tabular form. Using the tool, the list of most abundant chemical compounds present based on the highest intensity values can be known. This allows the determination of which appropriate options of treatment are possible in the case that the patient MS data tests positive for pancreatic cancer. The results of each application run can also be recorded due to the PDF export capability of the tool. This allows the user to keep track of past detections and analyses.

B. Project Dependencies

In order to run **PanCan Spectrum** smoothly, the following libraries are needed to be installed:

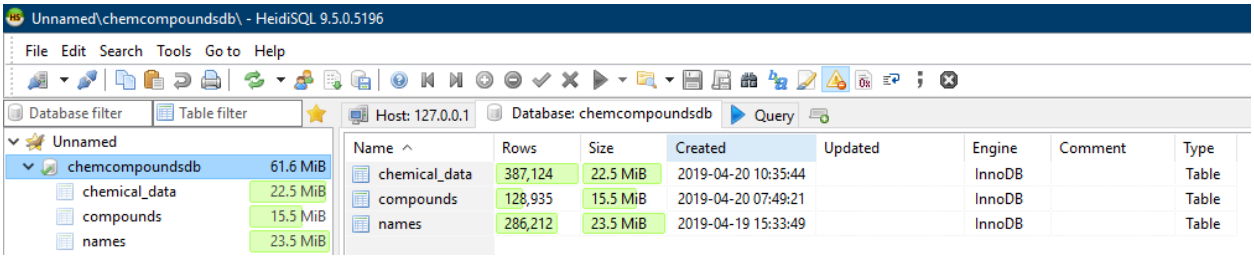
- SciPy
- NumPy
- sklearn
- pysimpleGUI
- pymzML
- mysql-client
- mysql-connector
- fpdf
- matplotlib
- pyplot

To install the libraries, run the command *'pip install <library-name>*' on a command terminal:



Example: Installing the pymzML library

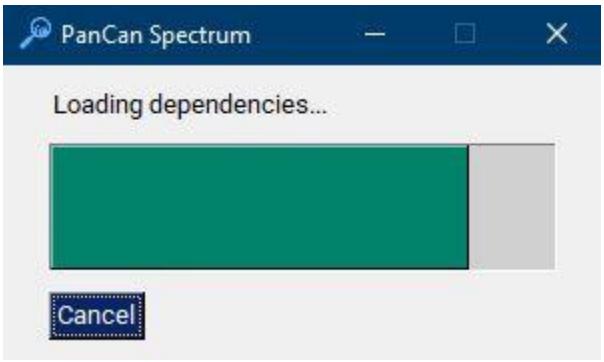
After installing the required libraries, import the chemical compounds database tables *compounds, names,* and *chemical_data* using the preferred MySQL administration tool:



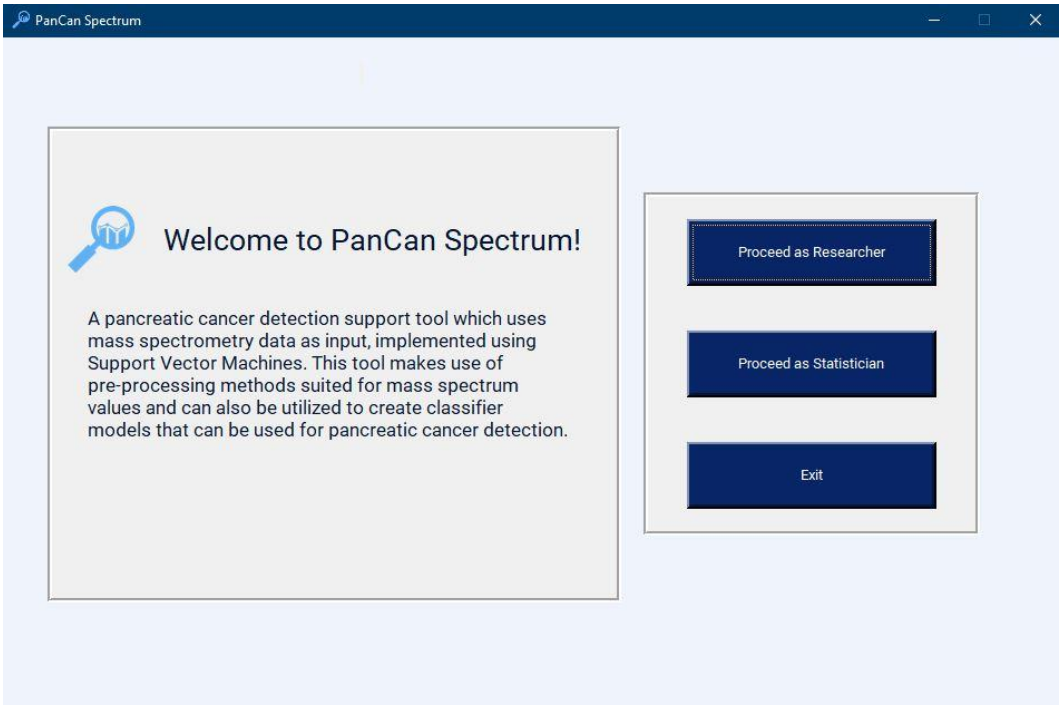
Also, ensure that the SVM models are located in a folder *svm_models* in the same location of the application. Afterwards, **PanCan Spectrum** can finally be used.

II. Starting the Tool

Clicking the **PanCan Spectrum** shows the following window. The user must wait until the application finishes loading.



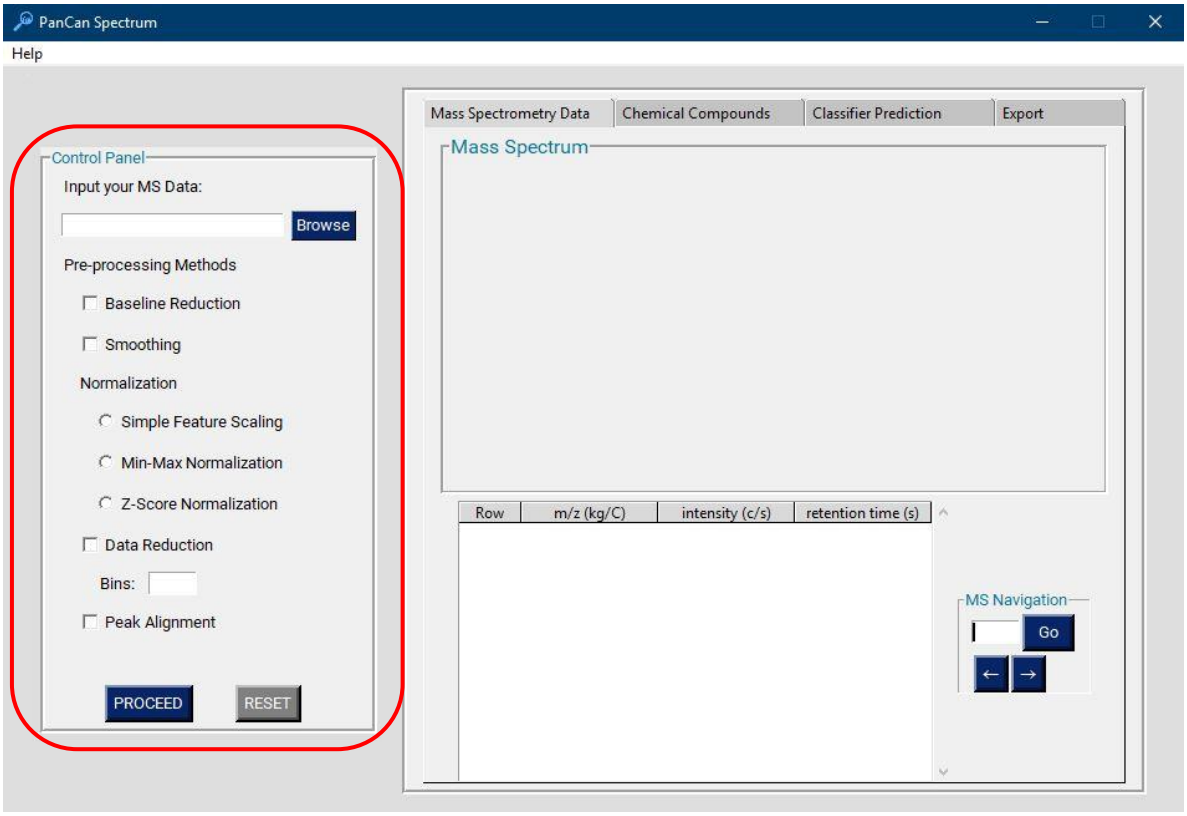
After **PanCan Spectrum** finishes loading, the following window is shown:



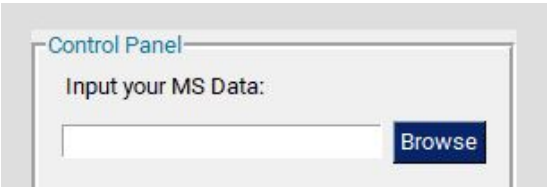
The user may choose to proceed using the tool as a **researcher**, or as an **administrator / statistician**. Most functionalities for each option would be different.

III. Using the Researcher Panel

A. Pre-Processing



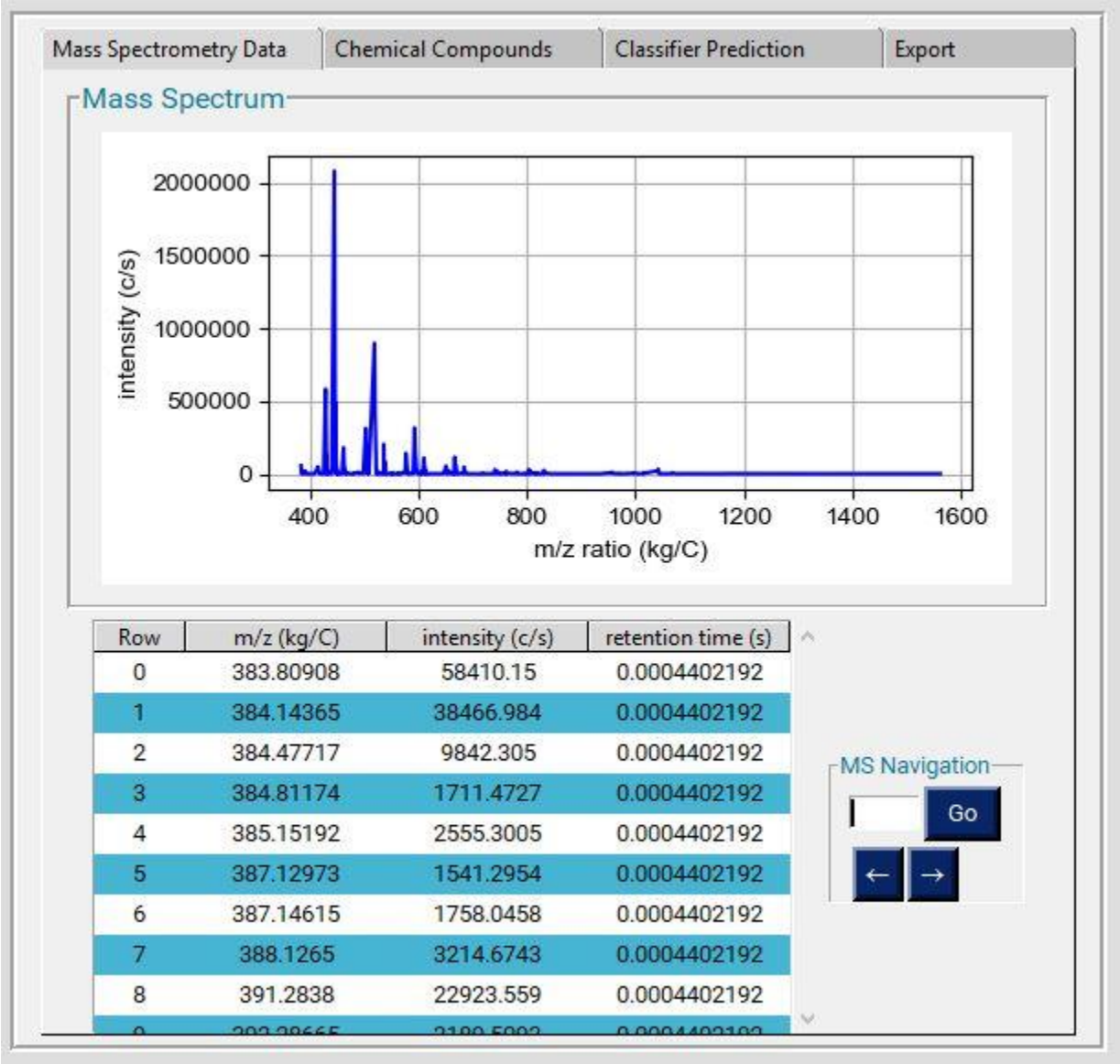
The control panel includes functionalities such as dataset browsing, data parsing, and all the pre-processing methods.



Simply click the browse button to select the *mzml file that you would be using as the input data.*

B. Data Visualization & Tabular Representation

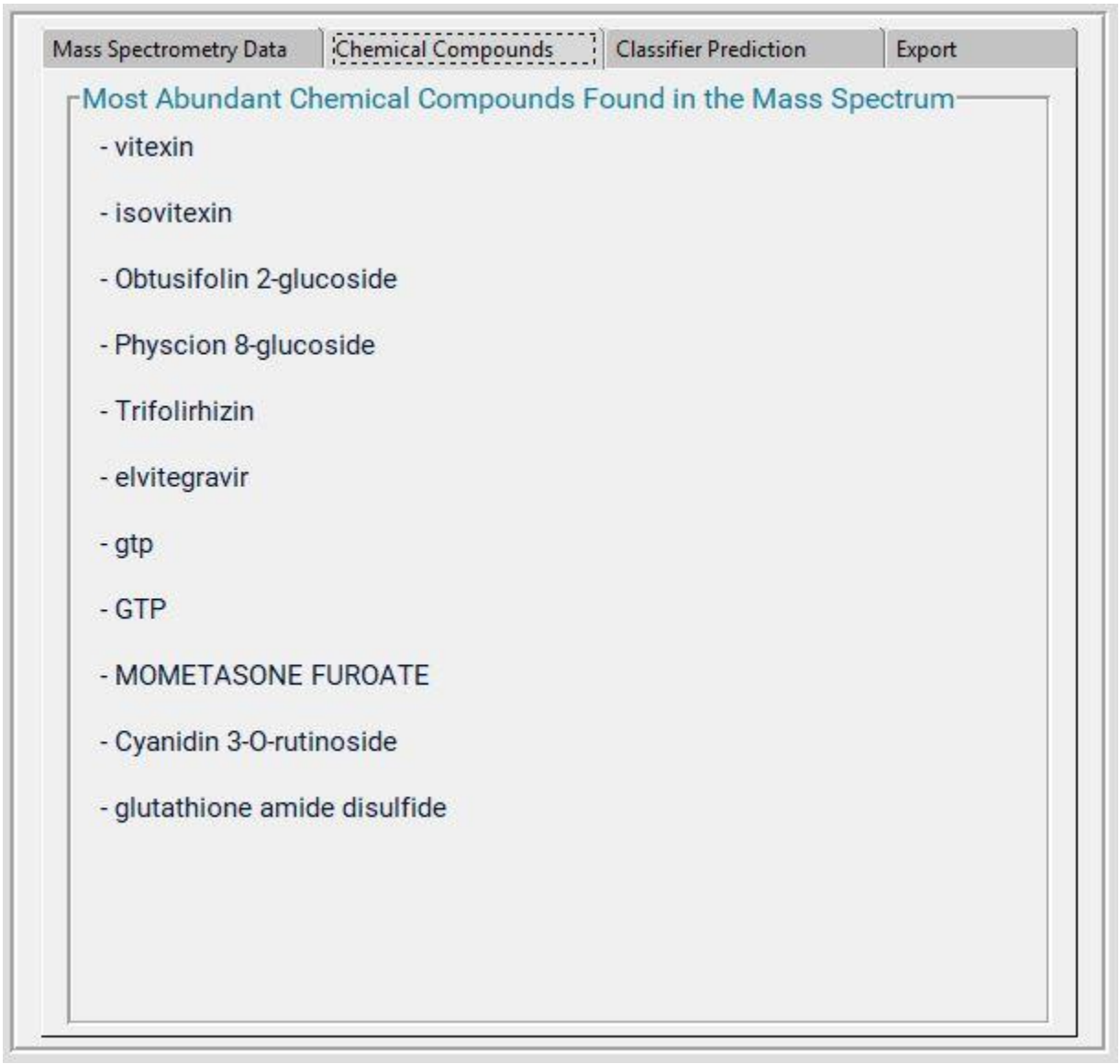
After selecting your preferred pre-processing methods, click **PROCEED**. If you choose to change your input dataset and/or pre-processing methods, click **RESET**. Once the user proceeds, the following would be displayed in the **Mass Spectrometry Data** tab:



Included are the mass spectrum, tabular data representation, and the MS data Navigation Panel. The user could use the navigator buttons or 'the input field + **GO**' option to browse through the spectra found in the inputted *mzml* file.

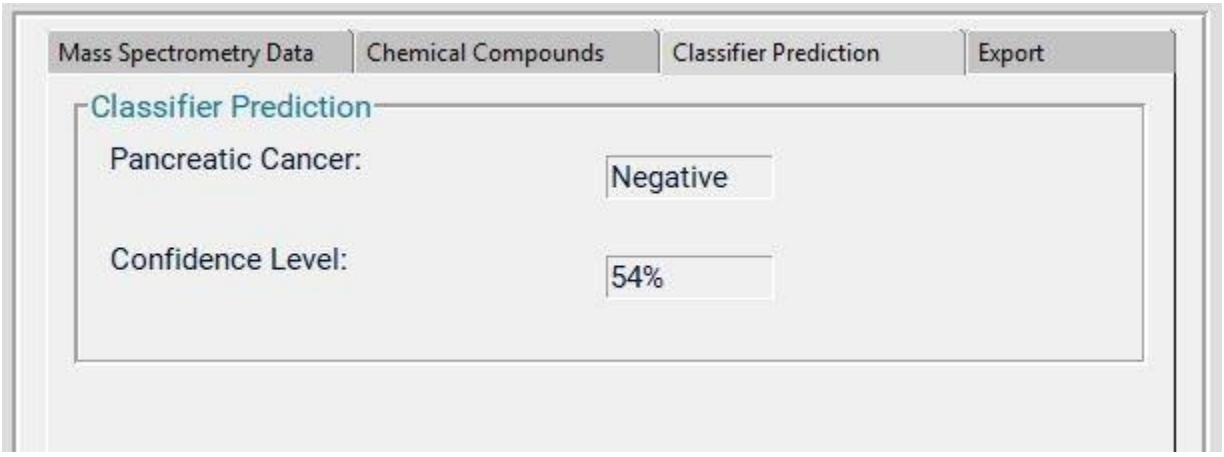
C. Listing Chemical Compounds

Also, after clicking **PROCEED**, the most abundant chemical compounds found in the current spectrum would be shown in the **Chemical Compounds** tab. The results here would be dependent on the mass spectrum shown in the first tab.



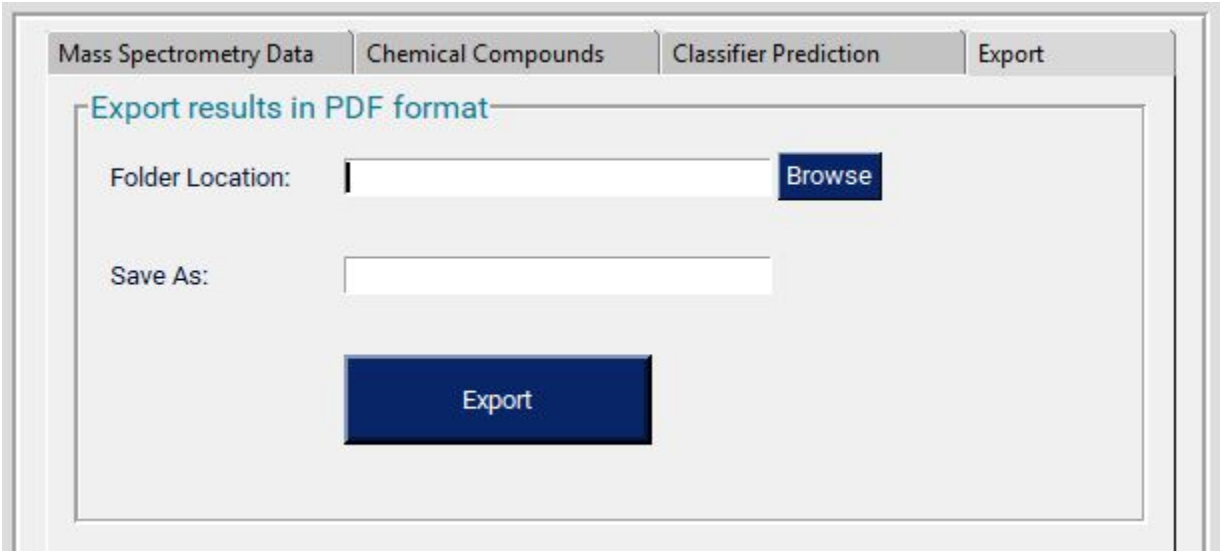
D. Prediction Results

The classification result is displayed in the **Classifier Prediction** tab. Included in the results are the prediction value (*Positive / Negative*), and the confidence rate of the prediction made by the tool.



E. Exporting Results

After performing the desired processing steps and browsing through the different tabs, the user may choose to export the results produced by tool using the **EXPORT** buttong in the **Export** tab.

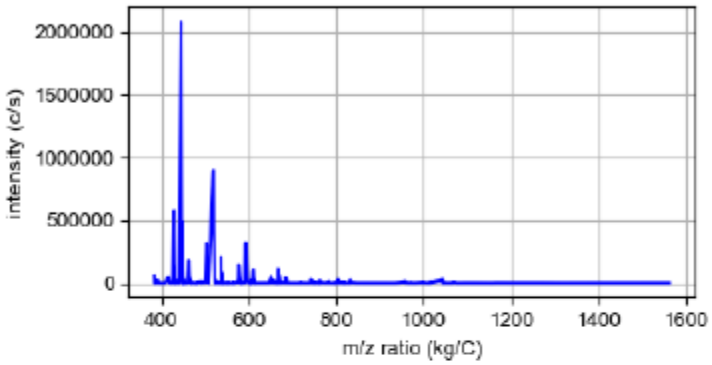


The user must input the download location and the filename of the PDF. The **Browse** button allows the user to select the folder to which the file would be saved.

After exporting the file, the user could now see the results in a PDF format (example below):

PanCan Spectrum
Results

Mass Spectrum (MS_A_1.mzML #1)



Prediction Results

Pancreatic Cancer: Negative

Confidence Level: 94%

List of Most Abundant Chemical Compounds

vitexin

isovitexin

Obtusifolin 2-glucoside

Physcion 8-glucoside

Trifolirhizin

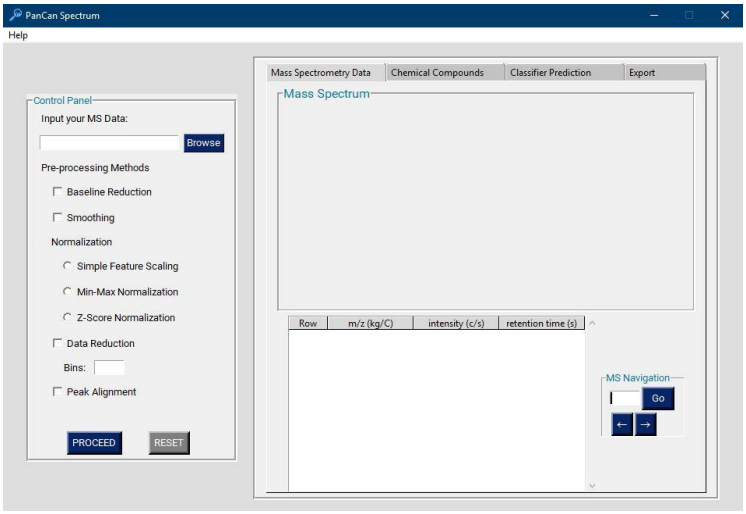
elvitegravir

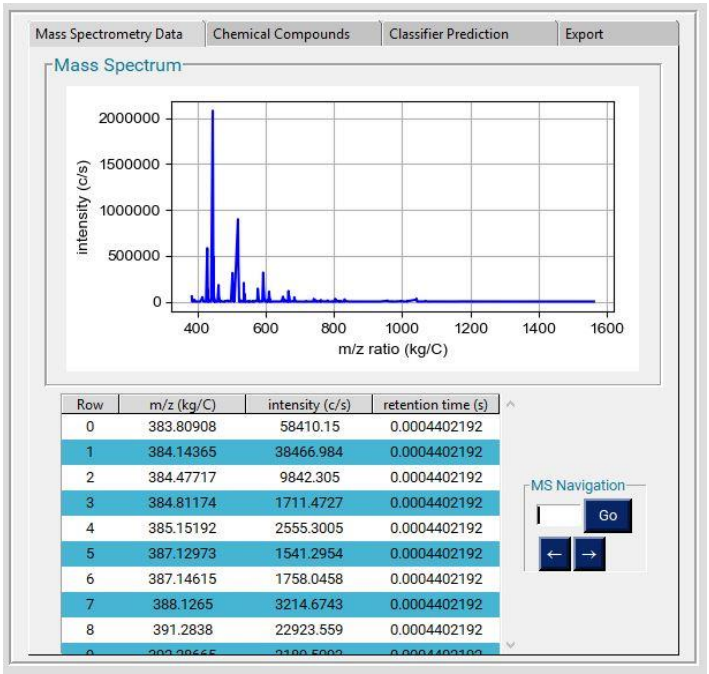
gtp

GTP

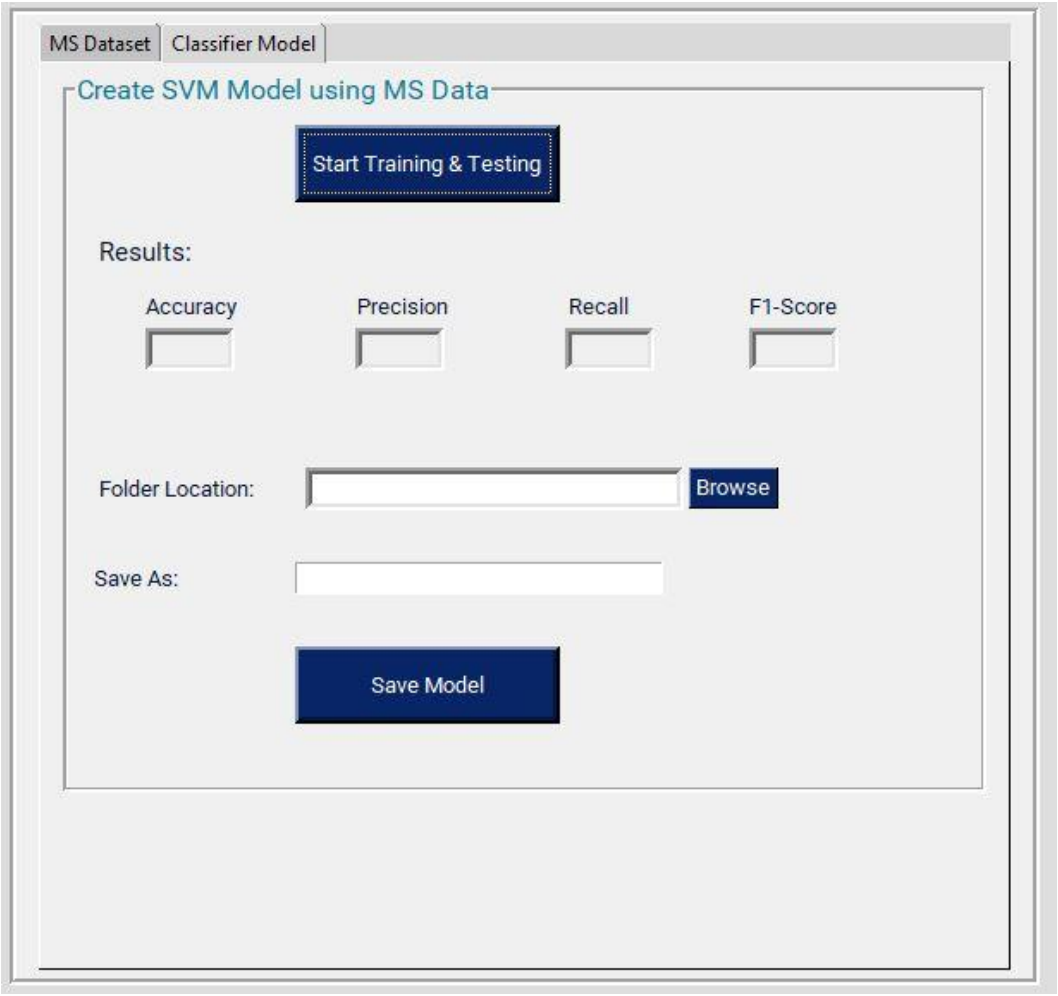
IV. Using the Statistician Panel

The statistician panel is similar to the researcher panel with regard to the pre-processing methods and mass spectrometry data visualization and tabular representation functionalities.





The statistician user may create a classifier model using the **Classifier Model** tab. After clicking the **Start Training & Testing** button, the data of the currently shown mass spectrum in the **Mass Spectrometry Data** would be used for creating an SVM model. Subsequently, the metric scores would be displayed.



After the creation of model, the user may input a download location for the model and specify the filename. Clicking the **Save Model** button would save the classifier model in the specified folder location, with the filename inputted.