User Manual - PLS analysis software

1 Details about the GUI of the PLS software

- The software mainly consists of two main tasks; partial least squares discriminant analysis (PLSDA) or classification and partial least squares regression (PLSR) analysis tasks.
- The menu bar consists of file, tasks and help
- The file menu enables the user to have access to the file system of the current operating system
- The tasks menu consists of the options for training and validation of the PLSDA and PLSR models
- The help menu consists of the details about the PLS analysis software.
- The main tool bar consists of the icons which act as the shortcuts for the execution of the commonly used tasks in the software
- From the left, the first icon opens the file system, the second icon closes the PLS analysis software, the third icon trains the PLSR model, the fourth icon trains the PLSDA model, the fifth icon validates the trained PLSR model, the sixth icon validates the trained PLSDA model, and the last shortcut icon displays details about the PLS analysis software.

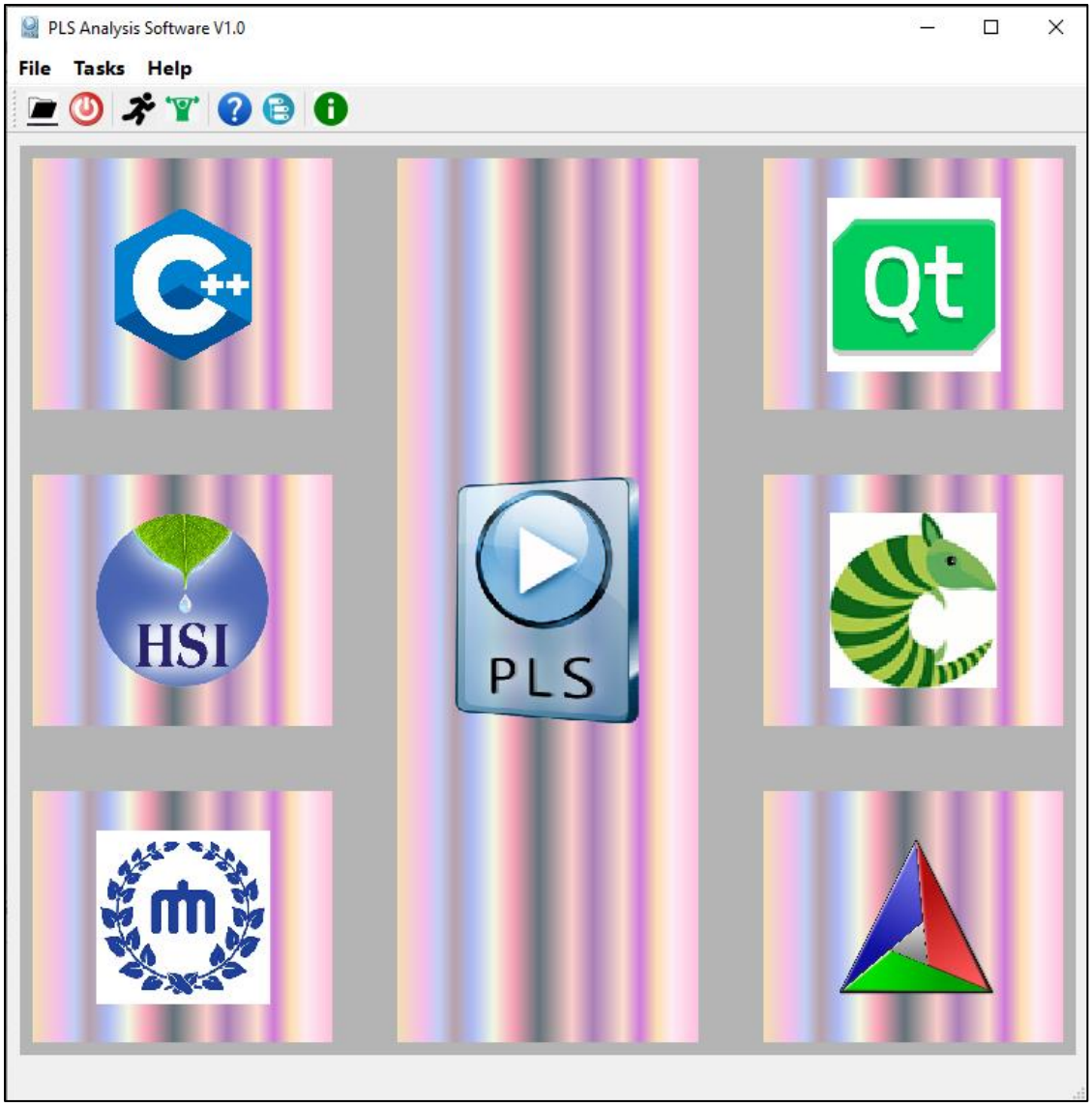

Figure 1: Home screen for the PLS analysis software

2 Train the PLSDA Model

- The spectra for normal and bloody eggs were organized in comma-separated values (.CSV) or text (.txt) file formats.
- The collected spectral data act as the predictor/independent/x variable which were used in the PLSDA model

- For the response/dependent/y variable, the value of zero (0) was assigned to the bloody eggs as the first class. The value of one (1) was assigned to the normal eggs as the second class.
- Both the predictor and the response variables were placed in one file and were used in the training of the PLSDA model.
- The data was randomly divided into training dataset (80 %) and test dataset (20%).
- The PLS analysis software was executed on the windows operating system by running the application (.exe) file.
- Within the software, the "train PLSDA model" icon was clicked, a window with parameters for training the PLSDA model displayed as illustrated in Figure 2.
- The spectral data was imported from file by clicking on the "Browse" button as illustrated in Figure 3. After import, the number of rows and columns for both the predictor and response variables were displayed in the software.
- The wavelength corresponding to the spectral data was loaded by clicking on the "Load Wavelength" button as illustrated in Figure 4.
- A preprocessing technique was selected and the preprocessed spectra was displayed in a dialog using the "Plot Spectra" button.
- The PLS analysis software provides the user with various preprocessing methods such as Standard normal variate (SNV), multiplicative scatter correction (MSC), Savitzky Golay 1 and 2, smoothing among others.
- The option of proceeding without any preprocessing was also provided. In this case, the raw spectra is used in the training.
- All the preprocessing methods were explored and the most suitable technique was chosen depending on the best accuracy obtained during training and testing.
- The "Maximum components" were specified. However, this act as initial number of PLS components. It is advisable to put more than 20 initial PLS components. The PLS analysis software is programmed to automatically calculate the optimum number of PLS components.
- A cross validation technique was chosen during training of the PLSDA model.
- The PLS analysis software provides cross validation techniques such as Leave-one-out cross validation also known as full validation, K-fold cross validation.

- The model was trained by clicking on “Train PLS-DA Model” button.
- The optimum number of PLS components were automatically calculated and displayed.
- The training plots such as predicted vs reference plot, the beta coefficient plot were viewed by clicking on the “View Classification Plots.”
- The beta coefficient was saved as a comma-separated values (.CSV) file.
- The beta coefficient was used in validation of the PLS-DA model
- The beta coefficient was loaded as the PLSDA model in the real time analysis software.

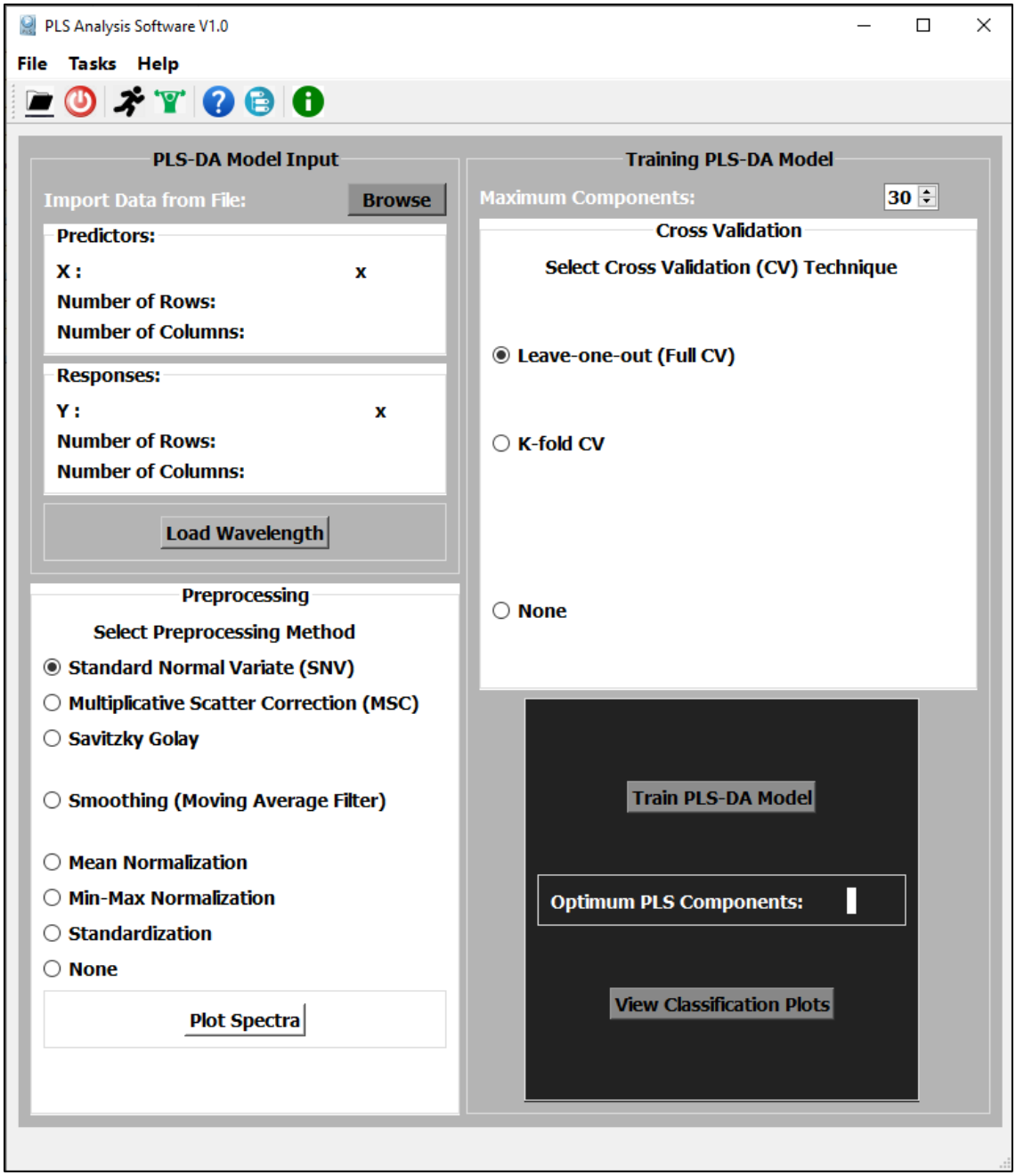

Figure 2: Parameters for training PLSDA Model

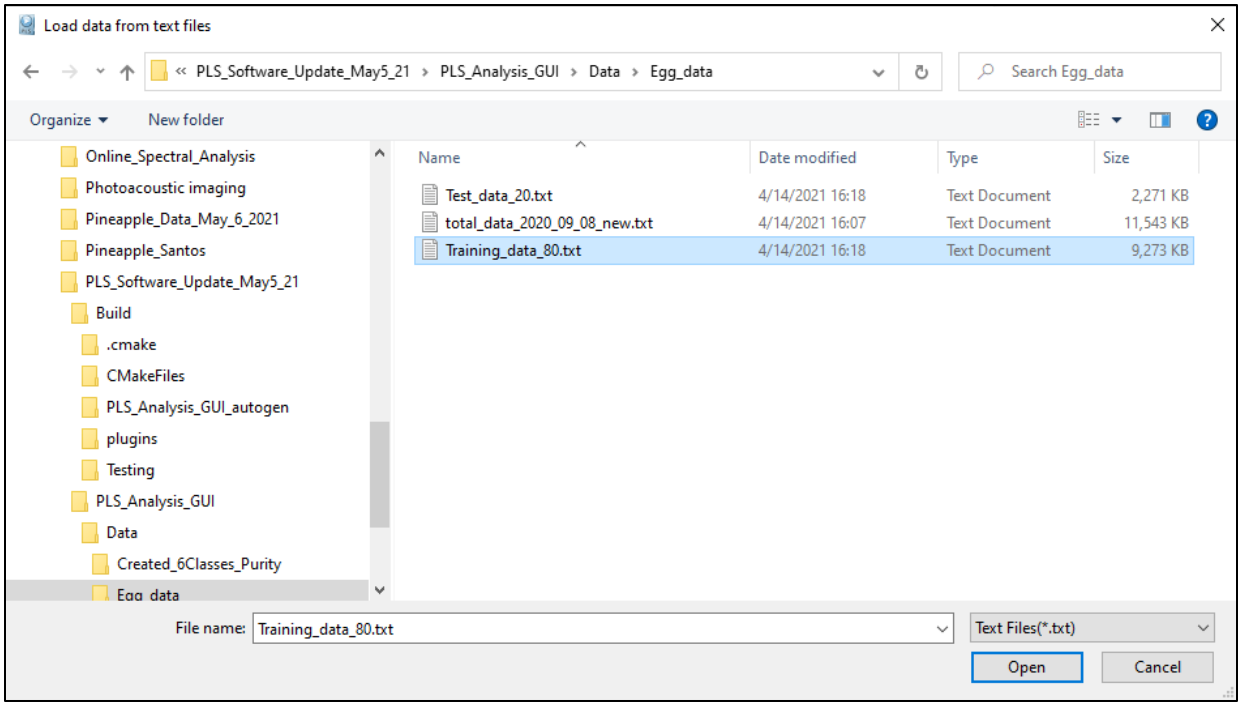

Figure 3: Loading the training dataset file from computer

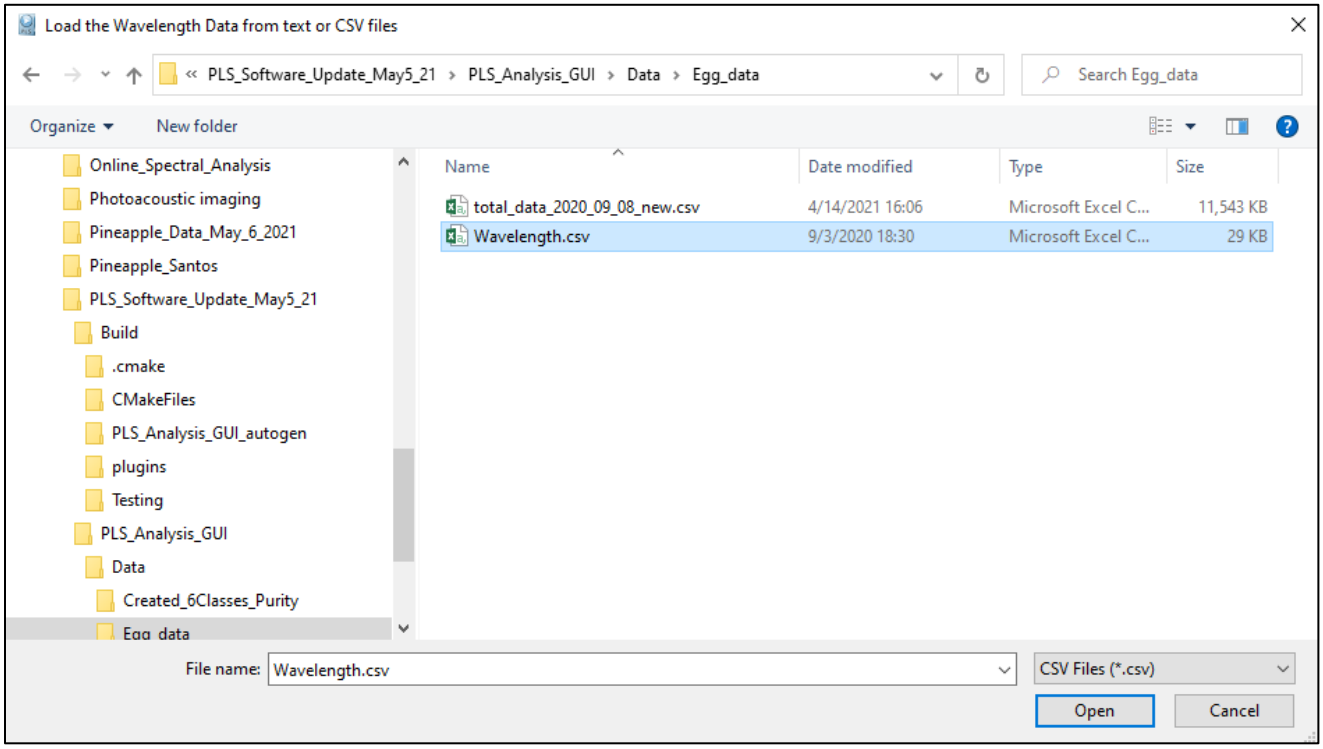

Figure 4: Loading the wavelength file from local computer

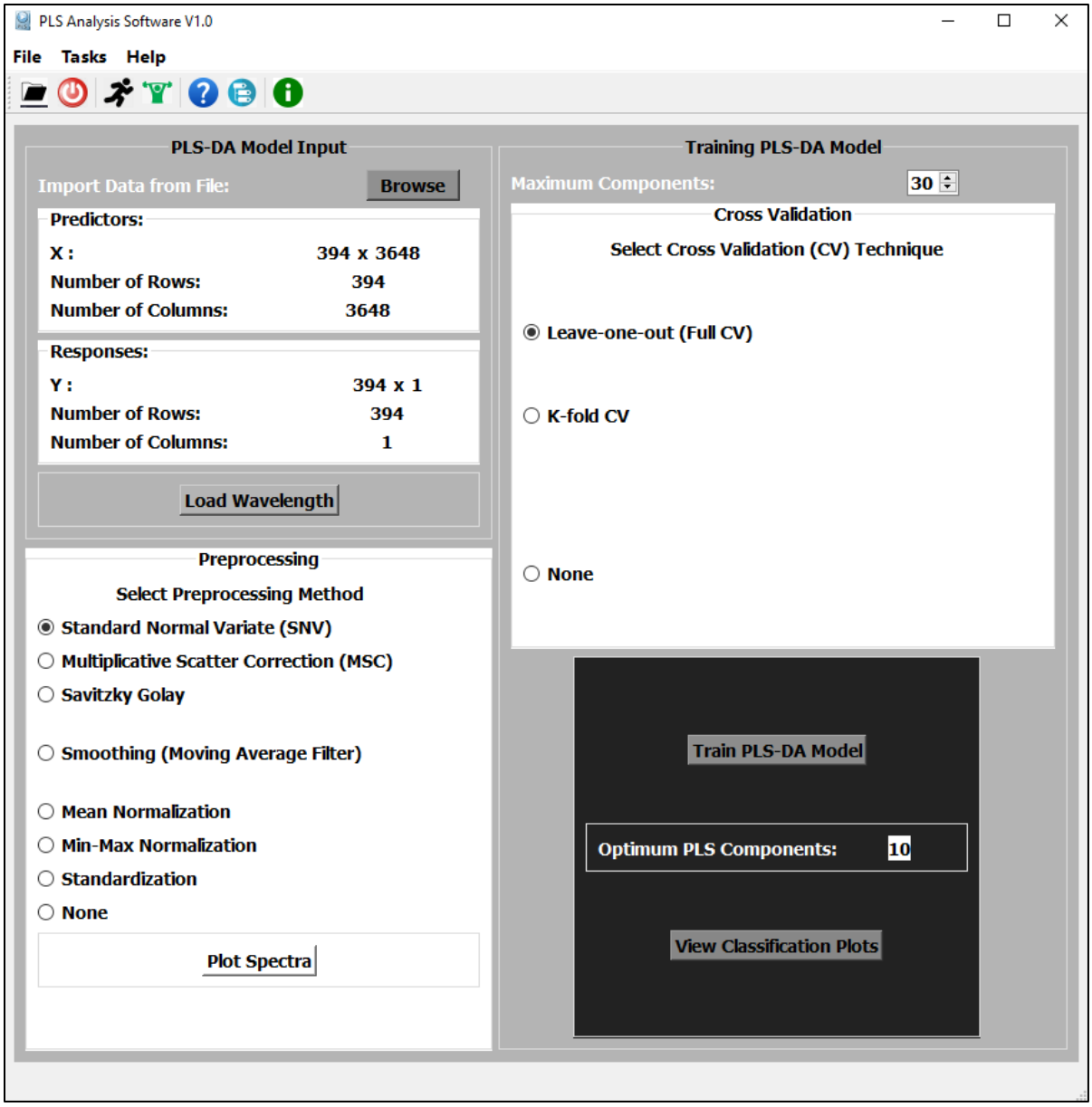

Figure 5: After Training the PLSDA model (SNV Preprocessing)

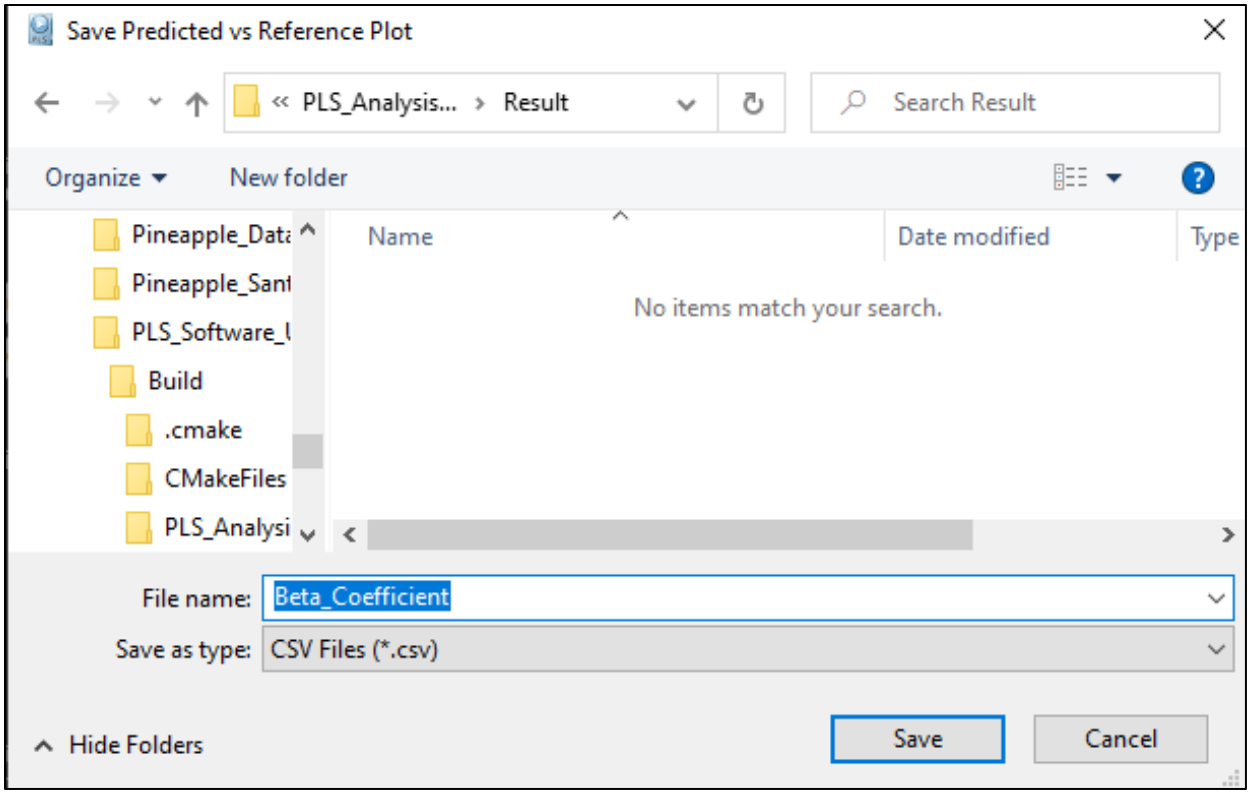

Figure 6: Save the beta coefficient to be used for online analysis

3 Validate the trained PLSDA Model

- The trained PLSDA model was validated by using the beta coefficient and the test dataset.
- The test dataset was imported from the local computer by clicking on the “File” button. After import, the number of rows and columns for both the predictor and response variables were displayed in the software as illustrated in Figure 8.
- The specific preprocessing technique which was used in the training of PLSDA model was chosen.
- The spectra of the test dataset was viewed using the “Plot Spectra” button.
- The model was validated using the “Validate PLS-DA Model” button.
- The validation plot was viewed by clicking on the “View Validated PLS-DA plots” button.
- The PLS analysis software was closed using the “Close application” shortcut icon.

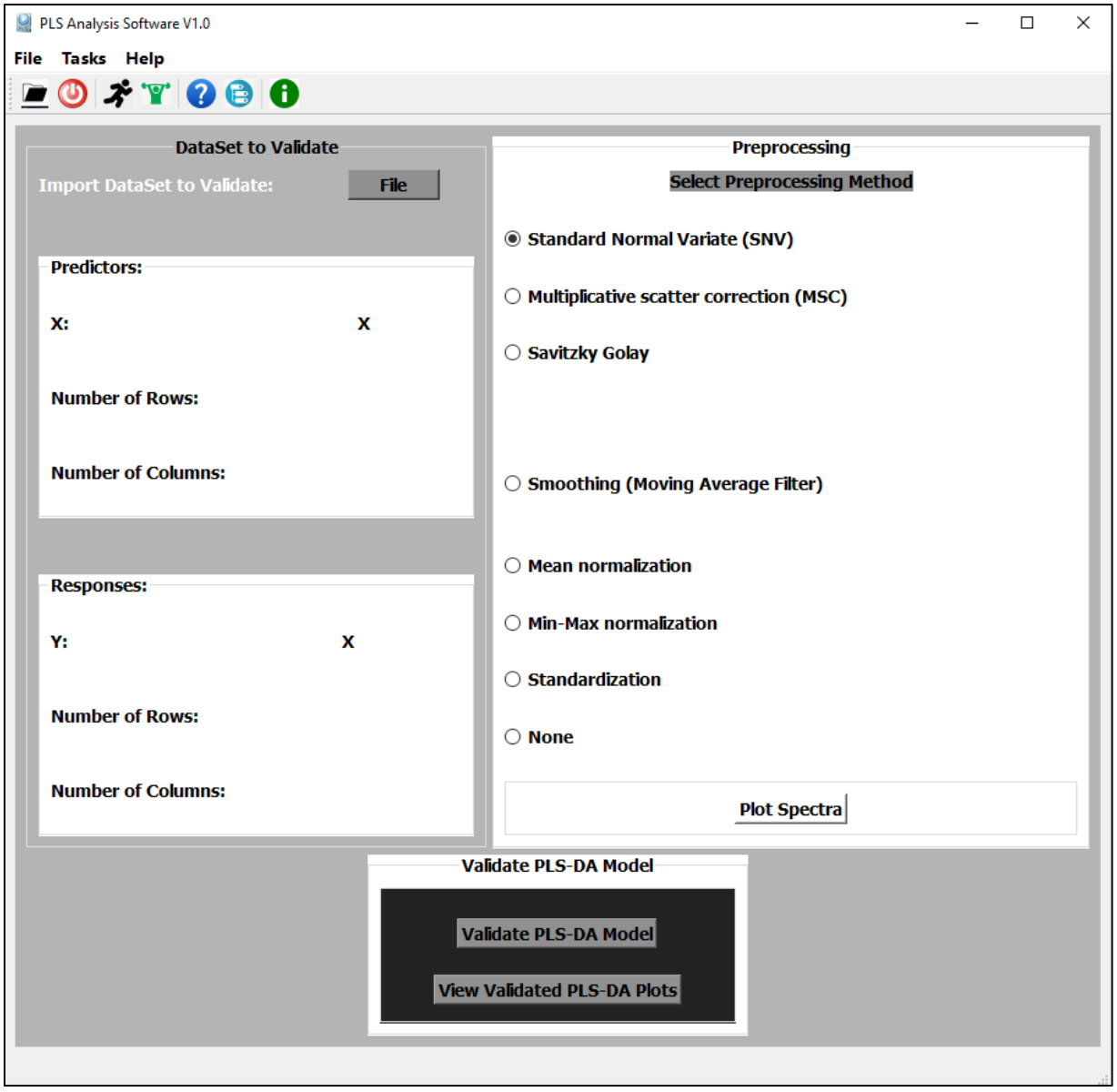

Figure 7: Window for validating the trained PLSDA model with a test dataset

Figure 8: After loading the test dataset for validation of trained PLSDA model