

Fast and Reliable Classification Analysis with Agilent MassHunter Classifier

Introduction

Food adulteration is a global challenge for the food industry. An adulterated food item can contain low-quality ingredients or lack certain vital constituents. These aspects can have serious health impact, and therefore, better analytical and data analysis procedures are required to detect food adulteration. Though an adulterated food can be identified using various techniques, one of the widely adapted technologies is classification analysis using metabolite fingerprints. Classification analysis involves recording the metabolite profile of authentic food items to create a prediction model. The model is then used to classify an unknown sample to a model group by matching the metabolite profiles. Besides food, classification analysis has also found use in other applications of similar requirements, such as flavors and fragrances.

The widespread adaptation of classification analysis benefits from empowering analysts to classify numerous unknown samples at the touch of a button. Broadly, an automated sample classification workflow requires sample preparation, detection, and data analysis to predict the unknown

samples, as shown in Figure 1. The data analysis step alone can take several weeks to months depending on the number and complexity of the samples. Agilent MassHunter Classifier, a simple classification tool, is Agilent's offering for fast and reliable detection of samples that are not authentic.

The overall workflow of MassHunter Classifier is as shown in Figure 2. It offers simplified sample import, in-depth data exploration and visualization options, as well as the ability to record results with a single click.

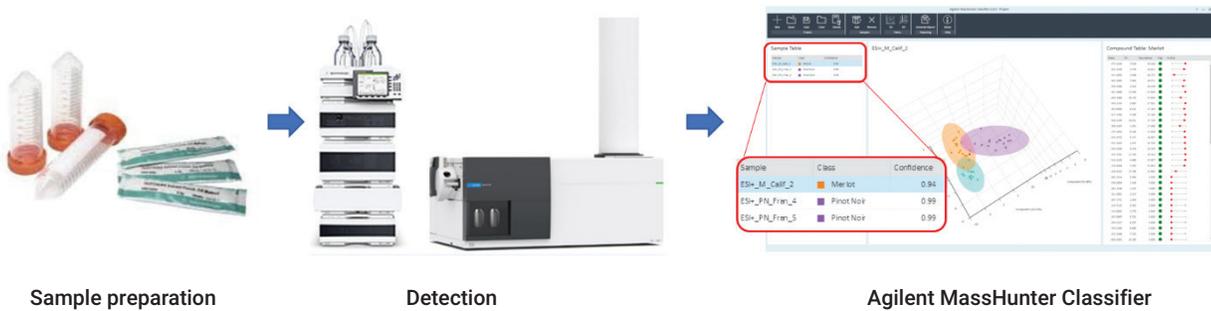


Figure 1. Sample classification workflow. It involves sample preparation, detection using mass spectrometry, and classification in Agilent MassHunter Classifier. The zoomed-in view shows classified samples along with confidence scores.

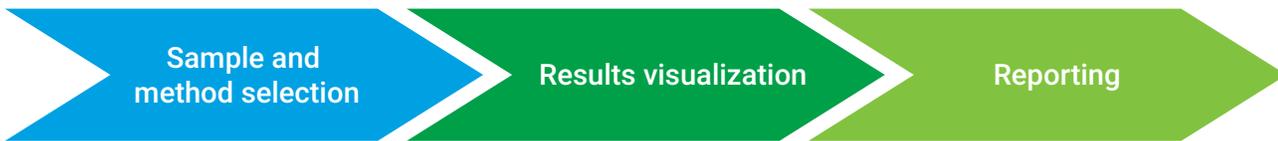


Figure 2. Agilent MassHunter Classifier workflow.

Using MassHunter Classifier

Agilent has a wide portfolio for separating and profiling metabolites in complex matrices (GC, LC, SQ, (Q)-TOF, etc.). The data collected from these platforms smoothly flow from the instrument into Agilent MassHunter Profinder for batch feature extraction followed by Agilent Mass Profiler Professional (MPP) for model building.

MassHunter Classifier uses a method and a model from Profinder and MPP for classifying the unknown samples. The analyst can add these files at the data import step as shown in Figure 3. One or more samples can be added or removed from the project later.

The tool quickly processes the samples and provides the user with a three-panel view to explore the results. Figure 4 shows the three panels, namely, Sample Table, Principal Component Analysis (PCA), and Compound Table.

Figure 3. Project creation window. This window allows import of method, model, and unknown sample(s). The analyst can work with either .d or .cef files for classification. When working with .d files, an Agilent Profinder method should be added.

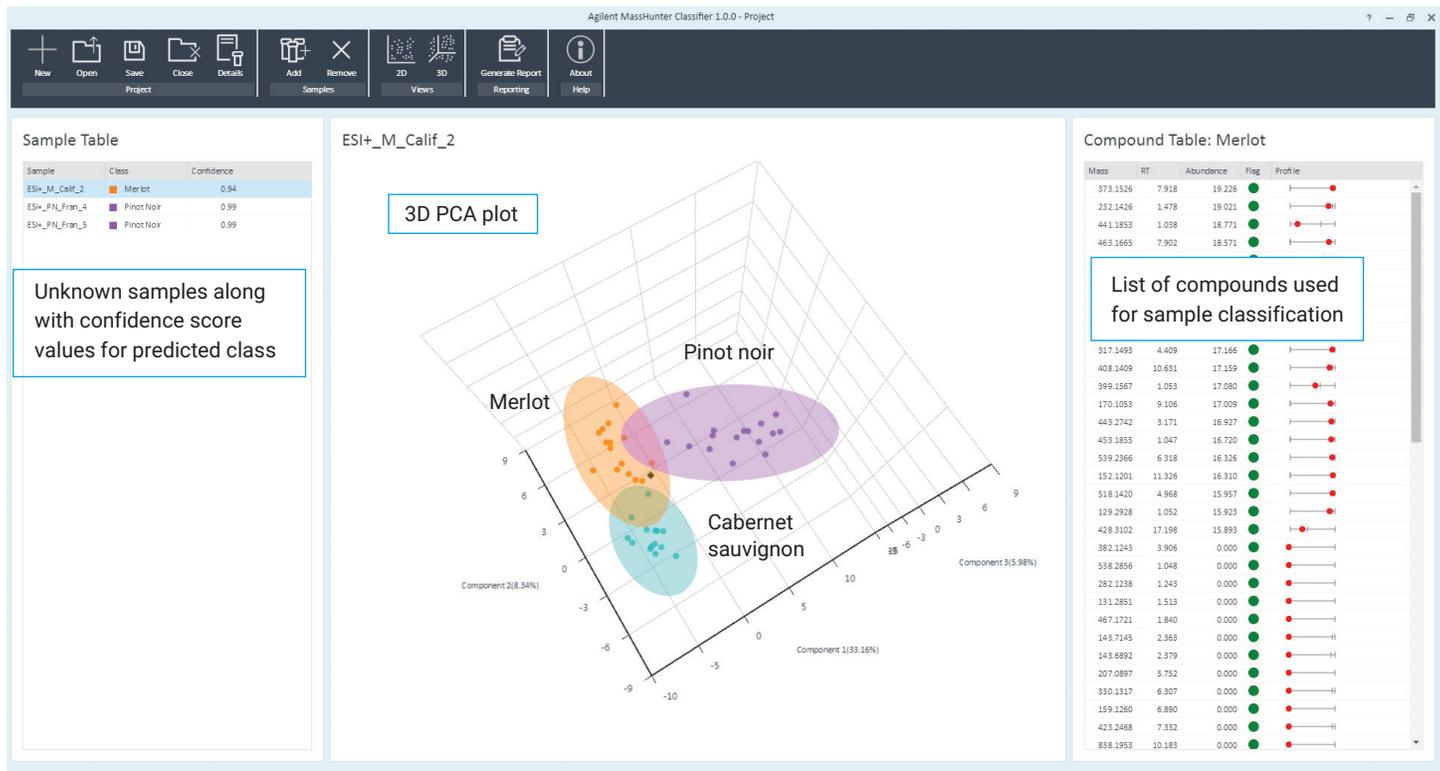


Figure 4. User interface of Agilent MassHunter Classifier showing wine varieties. Three unknown samples have been classified into Merlot and Pinot Noir classes.

Sample table

The unknown samples imported for classification are listed in the Sample Table. It shows the predicted class of the unknown sample and confidence score (Figure 5). A confidence score per class is provided for all class prediction models (random forest, linear discriminant analysis, partial least squares discrimination, support vector machine, Naive Bayes, decision tree, or neural network algorithm) except soft independent modeling of class analogy (SIMCA).

For prediction analysis run on models built with SIMCA, the distance of the unknown sample from the model is provided. This model also provides multiclass information for the classified samples. For example, if orange juice is adulterated with pineapple juice, these mixed samples can be classified using the model built with the SIMCA algorithm.

Principal component analysis (PCA)

A PCA plot, when launched from the toolbar, is available in the center panel. Figure 6 shows the 2D-PCA and 3D-PCA plots. The confidence ellipses are drawn from the model training data, and the unknown sample (selected from the

Sample Table) is represented as a black diamond. The circular dots on the PCA plot are training samples used for model building.

PCA plots are a great way to visualize the data distribution. However, overlapping areas in a PCA plot do not directly translate into classification of a sample. The classification of samples is based on the reported confidence score values.

Sample Table

Sample	Class	Confidence
ESI+_M_Calif_2	Merlot	0.94
ESI+_PN_Fran_4	Pinot Noir	0.99
ESI+_PN_Fran_5	Pinot Noir	0.99

Figure 5. Sample Table showing class and confidence score of classified samples.

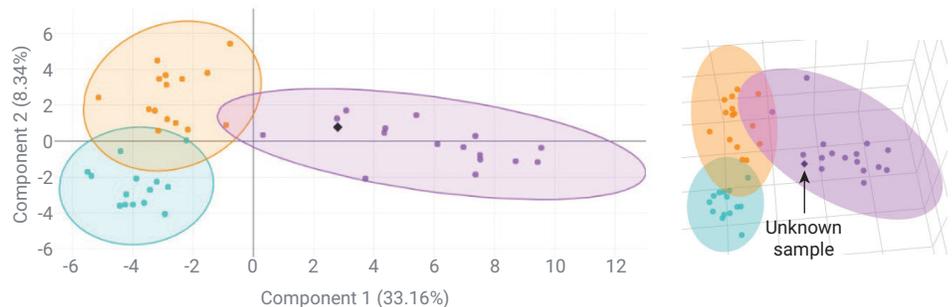


Figure 6. 2D-PCA (left) and 3D-PCA (right) graphs showing the unknown sample as a black diamond.

Compound table

The Compound Table lists the compounds that were used for the sample classification. In this example, Figure 7 shows the Compound Table from a wine study. The profile column shows the distribution of the compound abundance across the samples in the training set (samples used for the original model creation in MPP) for the predicted class. The red dot indicates the abundance of the compound in the unknown sample. If the abundance of the compound in the unknown sample lies within the abundance distribution of the compound in the training samples of the predicted class, the flag column has a circle filled green. Otherwise, the circle is filled white.

Compound Table: Pinot Noir

Mass	RT	Abundance	Flag	Profile
399.1567	1.053	17.368	●	
170.1053	9.106	17.304	●	
346.1243	2.804	17.251	●	
131.2851	1.513	17.064	●	
482.2112	1.076	16.946	●	
216.0630	2.813	16.868	●	
384.1774	11.015	16.801	●	
467.1721	1.840	16.797	●	
410.2525	1.889	16.745	●	
302.0430	11.415	16.592	●	
284.0897	8.071	16.582	●	
317.1493	4.409	16.566	●	
129.2928	1.052	16.348	○	
106.0421	5.080	16.189	●	
382.1243	3.906	16.006	●	

Figure 7. Compound Table showing flag information of compounds contributing to sample classification.

Reporting

MassHunter Classifier enables export of the classification results in a PDF report. The report includes information on the model used, samples tested, Sample Table, as well as a PCA plot. The user can choose to generate a report with all samples or select the samples of interest. An example report from MassHunter Classifier is shown in Figure 8.

A

Report

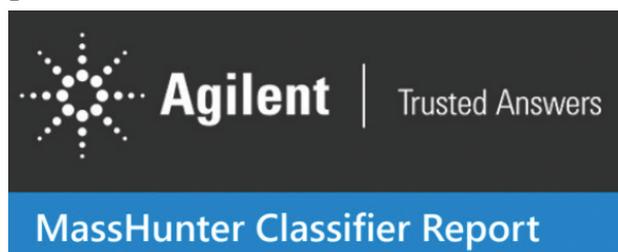
Title

Logo

Select Samples

- ESI+_M_Calif_2
- ESI+_PN_Fran_4
- ESI+_PN_Fran_5

B



Prediction model

MPP Experiment Name	Wine varieties
Prediction Model Name	Neural Network model on Type (Non-averaged)
Prediction Algorithm	Neural Network
Entities in Model	69
Creation Date	Tue Apr 30 19:20:26 IST 2019
Class Labels	Cabernet Sauvignon, Merlot, Pinot Noir
Model Accuracy	0.84

Samples

Sample	Class	Confidence
ESI+_M_Calif_2	Merlot	0.94
ESI+_PN_Fran_4	Pinot Noir	0.99
ESI+_PN_Fran_5	Pinot Noir	0.99

Figure 8. The report generation window to select samples (A) and a preview of the report showing prediction model details, classified samples, and PCA plot (B).

Conclusion

MassHunter Classifier software allows the classification of unknown samples in an automated workflow. The simplified workflow ensures easy adaptation of the tool for analysis in routine setup. The streamlined workflow for sample import, classification, and reporting addresses the major challenges of classification analysis.

www.agilent.com/chem

DE.6959722222

This information is subject to change without notice.

© Agilent Technologies, Inc. 2020
Printed in the USA, May 19, 2020
5994-1890EN

