



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils: A rapid untargeted chromatographic approach based on volatile compounds

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils: A rapid untargeted chromatographic approach based on volatile compounds / Rosa Palagano, Enrico Valli, Chiara Cevoli, Alessandra Bendini, Tullia Gallina Toschi. - In: LEBENSMITTEL-WISSENSCHAFT + TECHNOLOGIE. - ISSN 0023-6438. - STAMPA. - 130:(2020), pp. 109566-109572. [10.1016/j.lwt.2020.109566]

*Availability:*

This version is available at: <https://hdl.handle.net/11585/761698> since: 2020-11-06

*Published:*

DOI: <http://doi.org/10.1016/j.lwt.2020.109566>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

# Journal Pre-proof

Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils: A rapid untargeted chromatographic approach based on volatile compounds

Rosa Palagano, Enrico Valli, Chiara Cevoli, Alessandra Bendini, Tullia Gallina Toschi



PII: S0023-6438(20)30555-7

DOI: <https://doi.org/10.1016/j.lwt.2020.109566>

Reference: YFSTL 109566

To appear in: *LWT - Food Science and Technology*

Received Date: 27 December 2019

Revised Date: 10 April 2020

Accepted Date: 6 May 2020

Please cite this article as: Palagano, R., Valli, E., Cevoli, C., Bendini, A., Toschi, T.G., Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils: A rapid untargeted chromatographic approach based on volatile compounds, *LWT - Food Science and Technology* (2020), doi: <https://doi.org/10.1016/j.lwt.2020.109566>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

**CRedit Author statements**

Rosa Palagano: Investigation, Data curation, Writing - Original Draft.

Enrico Valli: Conceptualization, Writing- Reviewing and Editing

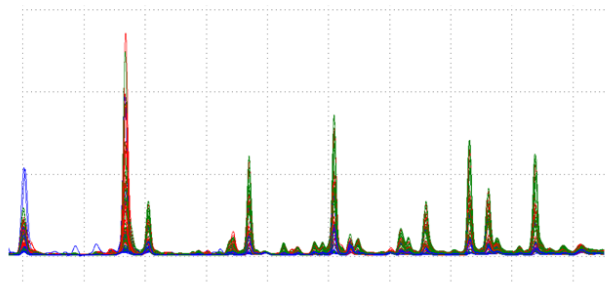
Chiara Cevoli: Methodology, Formal analysis, Visualization

Alessandra Bendini: Conceptualization, Writing- Reviewing and Editing

Tullia Gallina Toschi: Funding acquisition, Supervision

Journal Pre-proof

## 210 SAMPLES (EVOOs and VOOs)



## Extraction of raw data

**Data pre-treatment:**

1. Alignment of raw chromatograms (COW algorithm)
2. Mean centering
3. Unit vector normalisation



**PLS-DA**  
(Partial Least Square –  
Discriminant Analysis)



**ANN**  
(Artificial Neural Network)

Journal Pre

1 **Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils:**  
2 **a rapid untargeted chromatographic approach based on volatile compounds**

3  
4 Rosa Palagano<sup>a</sup>, Enrico Valli<sup>a</sup>, Chiara Cevoli<sup>a</sup>, Alessandra Bendini<sup>a,\*</sup>, Tullia Gallina Toschi<sup>a</sup>

5  
6 <sup>a</sup>Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Cesena, 47521,  
7 Italy

8  
9 \*Corresponding author: [alessandra.bendini@unibo.it](mailto:alessandra.bendini@unibo.it) (Alessandra Bendini), Department of  
10 Agricultural and Food Sciences (DISTAL) - University of Bologna, Piazza Goidanich 60, Cesena  
11 (FC), 47521, Italy, Tel +39 0547 338121.

**Abstract**

13 Many studies have shown that geographic origin is one of the most influencing factors in  
14 consumers' choice of olive oil. To avoid misleading consumers, European regulation has  
15 established specific rules to report the geographical origin of extra virgin (EVOOs) and virgin olive  
16 oils (VOOs) on the product label, even if an official analytical procedure to verify the origin has not  
17 been yet defined. In this work, a flash gas chromatography (FGC) untargeted approach based on  
18 volatile compounds, followed by a chemometric data analysis, is proposed for discrimination of  
19 EVOOs and VOOs according to their geographical origin (EU and extra-EU). A set of 210 samples  
20 was analyzed and two different classification techniques were used, one linear (Partial Least  
21 Square-Discriminant Analysis, PLS-DA) and one non-linear (Artificial Neural Network, ANN).  
22 The two models were also validated using an external data set. Satisfactory results were obtained  
23 for both chemometric approaches: with PLS-DA, 89% and 81% of EU and extra-EU samples,  
24 respectively, were correctly classified; for ANN, the percentages were 92% and 88%, respectively.  
25 These results confirm the reliability of the method as a rapid approach to discriminate EVOOs and  
26 VOOs according to their geographical provenance.

27

**Key words**

29 Virgin olive oil; Geographical origin; Chemometric analysis; Flash Gas Chromatography; Volatile  
30 compounds.

## 31 **1. Introduction**

32 Over the last 40 years many investigations have focused on understanding what attributes are  
33 important determinants in consumer choice, which have highlighted that geographic origin is one of  
34 the most influencing factors for olive oil (Dekhili, Sirieix, & Cohen, 2011; Del Giudice, Cavallo,  
35 Caracciolo, & Cicia, 2015). In order to ensure that consumers are not misled, the fourth article of  
36 the EU Reg. 29/2012 establishes that “*Extra virgin and virgin olive oil shall bear a designation of*  
37 *origin on the labelling*”. This means that for extra virgin (EVOOs) and virgin olive oils (VOOs)  
38 commercialized within the EU, it is mandatory to specify the geographical provenance of the  
39 product on the label following specific rules. If an oil comes from an EU Member State or third  
40 country, a reference to the EU Member State, to the EU, or to the third country must to be reported.  
41 In the case of blends of oils originating from more than one EU Member State or third country, one  
42 of the following phrases must be used: ‘blend of olive oils of European Union origin’ or a reference  
43 to the EU; ‘blend of olive oils not of European Union origin’ or a reference to origin outside the  
44 EU; ‘blend of olive oils of European Union origin and not of European Union origin’ or a reference  
45 to origin within the EU and outside the EU. An exception is the case where the olives were  
46 harvested in an EU Member State or third country other than that in which the mill where the oil  
47 was extracted is located. In this case, the designation of origin shall contain the following wording:  
48 ‘*(extra) virgin olive oil obtained in (the Union or the name of the Member State concerned) from*  
49 *olives harvested in (the Union or the name of the Member State or third country concerned)*’.

50 However, the regulation does not specify an official analytical procedure to verify the conformity of  
51 the label-declared geographical origin, and this has raised the interest of researchers to develop a  
52 reliable and effective method for the purposes of authentication (Conte et al., 2019). During recent  
53 years, different analytical techniques have been applied in order to find potentially useful markers  
54 and efficient instrumental approaches that are able to discriminate olive oils according to their  
55 geographical origin.

56 In this regard, traditional chromatographic techniques, analyzing both major and minor compounds  
57 either individually or in a combined way, coupled or not with specific statistical chemometric tools,  
58 have been investigated. A study in 2009 (García-González et al., 2009) proposed the application of  
59 artificial neural network (ANN) models for different levels of geographical classification (country,  
60 region, province, PDO) on a set of 687 EVOOs and VOOs from Spain, Italy, and Portugal, which  
61 were chemically characterized for the content of fatty acids, hydrocarbons, sterols, and alcohols.  
62 Other researchers have evaluated the triacylglycerol (TAG) content and composition to discriminate  
63 Moroccan oils (Bajoub et al., 2016) and Croatian samples (Peršurić, Saftić, Mašek, & Kraljević  
64 Pavelić, 2018). In addition, the stereospecific distribution of fatty acids in TAGs was reported to be  
65 useful in discriminating olive oils from different areas of North-Eastern Italy (Vichi, Pizzale, &  
66 Conte, 2007). Specific metabolites such as sterols and phenolic compounds have been investigated  
67 to identify the optimal markers, which may be a promising approach to discriminate oils according  
68 to geographical origin (Giacalone, Giuliano, Gulotta, Monfreda, & Presti, 2015; Ben Mohamed et  
69 al., 2018; Ghisoni et al., 2019). Interesting findings have also been recently reported on  
70 sesquiterpene hydrocarbons as geographical markers (Quintanilla-Casas et al., 2020). Moreover,  
71 volatile compounds have been amply studied by applying different instrumental techniques  
72 combined with chemometric data analysis (Kosma et al., 2017; Bajoub et al., 2018; Lukić, Carlin,  
73 Horvat, & Vrhovsek, 2019).

74 Rapid and innovative instrumental approaches have also been developed and tested to satisfy the  
75 need for simple, rapid, and environmentally friendly techniques (Valli et al., 2016). This critical  
76 review (Valli et al., 2016) overviews the principal applications of optical techniques (UV-Vis, NIR,  
77 MIR, RAMAN, NMR, and fluorescence spectroscopy), methods based on electrical characteristics,  
78 and instruments equipped with electronic chemical sensors (electronic nose and tongue) for  
79 discrimination of EVOOs and VOOs according to their geographical provenance. In addition to  
80 these approaches, other promising techniques include stable isotopes analysis (Angerosa et al.,  
81 1999; Chiochini, Portarena, Ciolfi, Brugnoli, & Lauteri, 2016; Bontempo et al., 2019), multi-



82 element fingerprint (Sayago, González-Domínguez, Beltrán, & Fernández-Recamales, 2018),  
83 differential scanning calorimetry (Mallamace et al., 2017), and GC-IMS (Gerhardt, Birkenmeier,  
84 Sanders, Rohn, & Weller, 2017).

85 Melucci and co-workers (Melucci et al., 2016) proposed the application of flash gas  
86 chromatography (Heracles II) and a multivariate data analysis to control the compliance of  
87 information on geographic origin declared in the label (“100% Italian” vs “non-100% Italian”) for  
88 the first time. This instrumental approach allows to carry out headspace analysis in a short time and  
89 the results are processed by chemometric tools following an untargeted approach. For this reason, it  
90 can be considered as a fingerprint method, since the data can be elaborated for sample classification  
91 that is not aimed towards identification and quantification of specific analytes. Following these  
92 preliminary results and the actual need for a rapid and effective method for geographical  
93 authentication of VOOs, the aim of this work was the application of FGC for rapid discrimination  
94 of 210 EVOOs and VOOs according to geographical provenance. In this case, the categories  
95 considered for samples classification were EU member states vs extra-EU countries, and the data  
96 obtained were analyzed by applying two different classification techniques, one linear (Partial Least  
97 Square-Discriminant Analysis, PLS-DA) and one non-linear (Artificial Neural Network, ANN).

98

## 99 **2. Materials and methods**

### 100 **2.1 Samples**

101 A total of 210 EVOOs and VOOs with a different geographical origin were collected directly from  
102 companies that were also asked to provide, when available, information about location of the mill,  
103 type of plant used, olive variety, and commercial category (Table S1, Supplementary material).  
104 Samples were collected during three crop years: 84 samples were collected during the crop season  
105 2016-2017, 90 during the crop season 2017-2018, and 36 during the crop season 2018-2019.  
106 Samples belonging to both the EVOO and VOOs commercial categories were included in this

107 study, considering that the indication of the geographical origin on the product label is mandatory  
108 for both these grades.

109 According to geographical provenance, samples were distributed in 3 classes (Table 1): “EU” for  
110 oils coming from EU member states; “extra-EU” for oils coming from extra-EU countries (outside  
111 EU); “blends” for samples obtained by mixing oils coming from different EU state members or oils  
112 coming from EU state members and third countries.

113 Aliquots of each sample (50 mL) were stored at -18 °C in dark plastic bottles.

114 Oils were defrosted for at least 12 h and stored at 12°C before analysis.

115

## 116 **2.2 Volatile compounds analysis by FGC**

117 The analysis of volatile compounds was carried out using the FGC Electronic Nose Heracles II  
118 (Alpha MOS, Toulouse, France). The instrument was equipped with two metal capillary columns  
119 (MXT-5: 5% diphenyl, 95% methylpolysiloxane; MXT-1701: 14% cyanopropylphenyl, 86%  
120 methylpolysiloxane, for both columns: 10 m length, 180 µm internal diameter, 0.4 µm film  
121 thickness) working in parallel mode with a different polarity of the stationary phase. This permits  
122 slight differences in the separation capability of molecules detected by a FID applied at the end of  
123 each column.

124 Each sample was analyzed in triplicate, weighing  $2 \pm 0.1$  g of oil in a 20 mL vial sealed with a  
125 magnetic plug. For analysis, the vial was placed in a shaker oven for 20 min at 40 °C and 500 rpm.  
126 Next, 5 mL of the headspace was collected, introduced in a splitless injector (injector temperature  
127 200 °C, injection speed 100 µL/s, carrier gas flow, to ensure a fast transfer of the sample from the  
128 inlet to the trap, 30 mL/min), and adsorbed on a Tenax<sup>®</sup> TA trap maintained at 40 °C for 60 s to  
129 concentrate the analytes. The syringe temperature was set at 70 °C. Subsequently, desorption was  
130 obtained by increasing the trap temperature to 240 °C in 93 s and the sample was injected (pressure  
131 of the carrier gas at the column head was 40 kPa.) and split (split flow 5 mL/min) into the two  
132 columns. The thermal program started at 40 °C (held for 2 s), increased up to 80 °C at 1 K/s, and

133 then to 250 °C at 3 K/s. Hydrogen was used as the carrier gas with a pressure from 40 kPa to 64  
134 kPa, increasing with a rate of 0.2 kPa/s. At the end of each column, a FID detector (detector  
135 temperature 260 °C) was placed and the acquired signal was digitalized every 0.01 s. The software  
136 used to control the instrument was AlphaSoft version 14.5.

137

### 138 **2.3 Data processing**

139 For data analysis, the full chromatograms were processed by applying chemometric data analysis  
140 with an untargeted approach. The raw data from each chromatogram (intensity values for each point  
141 of the chromatogram considering that the signal was digitalized every 0.01 s, Palagano et al., 2019  
142 [Dataset]) were exported from the software of the instrument and the data set with all the samples  
143 was imported into MatlabR2018a<sup>®</sup>. As data pre-treatment, chromatograms were aligned by COW  
144 (Correlation Optimized Warping) algorithm (Tomasi, Van Den Berg, & Andersson, 2004) and  
145 autoscaled (mean-centering followed by division of each column (variable) by the standard  
146 deviation of that column). Preliminary tests showed that chromatograms obtained from the MXT-5  
147 column had a discriminant power higher than the other one (MXT-1701) and for this reason the  
148 classification models were developed considering only this column. Considering the reduced  
149 number of samples for the classes “blend EU” and “blend EU/extra EU”, these oils were grouped  
150 together with “EU” and “extra-EU” samples, respectively. While this can be considered to be  
151 simplification, the decision was based on the fact that the percentage of extra-EU oils was higher  
152 than EU ones in these EU/extra-EU mixtures. This means that for the data elaboration only two  
153 sample categories were considered: “EU” and “extra-EU”.

154 Two different statistical techniques were used to classify samples according to geographical origin,  
155 the first (PLS-DA) based on a linear approach, and the second (ANN) on a non-linear approach.

156 In particular, the PLS-DA model was built using the PLS Toolbox for Matlab2018a<sup>®</sup>: intensity  
157 values of each point of the chromatogram, for a total of 19,900 data points, were used as variable X  
158 (matrix X), while the origin (“EU” and “extra-EU”) was implemented as variable Y (binary

159 variables, 0 - 1). The sample data set was split into a calibration/full-cross validation set (75% of the  
160 sample) and an external validation set (25% of the samples) using the Kennard-Stone method  
161 (selects samples that best span the same range as the original data, but with an even distribution of  
162 samples across the same range) (Daszykowski, Walczak, & Massart, 2002). The threshold value  
163 useful to define the category of each sample was defined using a probabilistic approach based on  
164 Bayes's rule.

165 The ANN model was performed by using the Neural Net Pattern Recognition tool for  
166 Matlab2018a<sup>®</sup>. Specifically, a Multi-Layer Perceptron (MLP) neural network was built to predict  
167 the specific class to which samples belong using a non-linear method. For input and hidden layers,  
168 linear and logistic activation functions, respectively, were used, while for output layer the SoftMax  
169 function was applied. From a statistical point of view, with the SoftMax activation function and  
170 cross entropy error, the output is interpretable as posterior probabilities for categorical target  
171 variables (Bishop, 1995). One nominal output variable is returned, assuming that the target output is  
172 1.0 in the correct class output, and 0.0 in the non-correct class. Looking for the best classification  
173 ability, different node numbers in the hidden layer and combinations were tested. The convergence  
174 of ANN was ruled by a back propagation algorithm.

175 The original data set was divided into training (55%), validation (20%), and external validation test  
176 (25%) sets. The latter was composed of the same samples used for the external validation of the  
177 PLS-DA model and selected using the Kennard-Stone method. The training set was used to  
178 calculate the transfer function parameters of the network, the validation set to measure network  
179 generalization and to halt training when generalization stops improving, and the test set was treated  
180 as an unknown, the correct classification of which indicates that the neural network is performing  
181 well.

182

### 183 **3. Results and discussion**

184 A set of 210 EVOOs and VOOs were analyzed for their volatile profile by FGC. Considering the  
185 large amount of data and aim of this work, chemometric data analysis following an untargeted  
186 approach was carried out.

187 For elaborations, samples were grouped into two categories: “EU”, which included oils from single  
188 EU state members and blends of oils from different EU countries, and “extra-EU” that consisted of  
189 oils from single countries outside the European Union and blends of oils from the EU and third  
190 countries.

191 In Fig. 1-a the mean chromatogram of “EU” and “extra-EU” categories, obtained averaging the  
192 intensity of each variable for all “EU” or “extra-EU” samples, is reported: even if almost all peaks  
193 are concentrated in the initial part of the chromatogram (between 2000 and 10000 variables), a clear  
194 difference, in terms of variable intensities, exists between the two groups, thus confirming the  
195 discriminating power of the volatile profile with respect to the geographical origin (Melucci et al.,  
196 2016; Lukić, Carlin, Horvat, & Vrhovsek, 2019).

197 Concerning the PLS-DA results (4 latent variables), the values of the estimated Y variable  
198 (geographical category) by the model in cross and external validations are shown in Fig. 2. The  
199 dotted line identifies the threshold value used to define the attribution of samples to different  
200 classes. Regarding the location of each sample, a greater distance from the threshold line can be  
201 interpreted as a better classification capacity of the model.

202 The results, in terms of percentage and number of samples correctly classified, are reported in Table  
203 2. The percentage ranged from 80.8% to 91.2%. The values obtained for the “EU” category were  
204 higher, likely because of the greater number and variability of samples used to build the model. The  
205 external validation percentages were lower compared to those obtained for the cross-validation as  
206 expected, but the results can be considered more robust since they were obtained considering the  
207 25% of samples that were not used to build the model.

208 The ROC (Receiver Operating Characteristic) curve, reported in Fig. 3, summarizes the trade-off  
209 between sensitivity (number of samples predicted to be in the class divided by number actually in

210 the class) and specificity (number of samples predicted to not be in the class divided by the actual  
211 number not in the class) for the PLS-DA classification model (external data set). The area under the  
212 curve (AUC=0.9365) and the height values of sensitivity (0.89) and specificity (0.8) suggest that the  
213 model was characterized by a good degree of discrimination. The specificity and sensitivity as  
214 threshold variation is reported on the right. The dotted line identifies the threshold value used in the  
215 model and is shown in Fig. 2.

216 The VIP (Variable Importance in Projection) score obtained by the PLS-DA confirmed that the  
217 section of the chromatogram ranging from 2000 to 10,000 variables has a major contribution to  
218 sample discrimination (VIP values greater than 1) according to geographical origin (Fig. 1-b).

219 Focusing on incorrectly classified samples, a specific trend as a function of characteristics that  
220 could usually affect the volatile profile of the oil (such as the commercial category, olive cultivar, or  
221 specific country of origin) was not seen, thus not justifying the observed misclassifications.

222 Results related to the probabilistic approach are shown in Fig. 4. The graph refers to the category  
223 “EU”: this means that the higher a sample is located, the higher the probability for which it is  
224 classified as member of the “EU” category. As a consequence, oils classified as members of the  
225 other category (extra-EU) are located in the bottom area of the graph. In this case, the threshold  
226 value is fixed at 0.5, corresponding to a probability of 50%: a sample classified with a probability  
227 lower than this is considered as not correctly grouped. It is also interesting to note that most samples  
228 were correctly classified with a probability between 90% and 100%.

229 Regarding ANN, an early stopping technique was used to select the number of training cycles  
230 (epochs) to avoid over-fitting, using the validation set to monitor the prediction error. An example  
231 of this procedure is reported in Fig. 5, where the best ANN training was characterized by 47 epochs.  
232 Above this point, the error increased further indicating that the ANN tends to overfit. Consequently,  
233 the results of ANN are related to these iterations.

234 The best prediction results were obtained with a three layers network (input, hidden and output),  
235 having 5 nodes in the hidden one; a larger number of nodes did not increase the network  
236 performance.

237 The classification results, in terms of number and percentage of samples correctly classified  
238 (according to the highest class probability), are summarized in Table 3, for the validation and  
239 external data sets. The external data set was composed of the same samples selected to validate the  
240 PLS-DA model. As reported for the PLS-DA model, even in this case higher percentages (from  
241 92.3% to 97.5%) were achieved for the “EU” category in all three data sets.

242 In particular, comparing the results of the external validation data (the same for both classification  
243 methods), it is possible to note that higher percentages were obtained using the ANN for both the  
244 “EU” and “extra-EU” categories. In particular, an increment of 3.8% and 7.6% of correctly  
245 classified samples was obtained, respectively. The ROC curve (specificity vs sensitivity) and AUC  
246 value (0.9526) of the ANN model (external data set) confirms the higher classification power of this  
247 techniques compared to PLS-DA (Fig. 6). This is probably due to the fact that the ANN model is  
248 based on a non-linear approach.

249 In general, the percentages obtained were slightly lower than those reported by other studies based  
250 on volatile compounds and chemometric untargeted data analysis (Gerhardt, Birkenmeier, Sanders,  
251 Rohn, & Weller, 2017; Bajoub et al., 2018; Lukić, Carlin, Horvat, & Vrhovsek, 2019). This aspect  
252 can be explained by the large variability, in terms of geographical origin, olive variety, and  
253 commercial category of the samples analyzed, which represents a strong point of this work.

254 The results described herein confirm the suitability of FGC for verifying geographical traceability  
255 of EVOOs and VOOs, even using untargeted chromatographic signals of the volatile fraction as  
256 variables for multivariate analysis (Melucci et al., 2016). An in-house validation of this analytical  
257 method, performed to verify that a repeatable and reproducible signal, with sufficient sensitivity to  
258 collect the valuable information from the samples, has been carried out which underlined the good  
259 performance of the technique; this will be discussed in more detail in a subsequent publication.

260

261 **4. Conclusions**

262 In this work, the application of FGC for volatile compounds analysis combined with untargeted  
263 chemometric data analysis (PLS-DA and ANN) to discriminate EVOOs and VOOs with different  
264 geographical origin was presented. For both elaborations, satisfactory results, in terms of  
265 percentages of correctly classified samples, were obtained: PLS-DA (external validation) allowed  
266 classification of 89% and 81% of “EU” and “extra-EU” samples, respectively; for ANN (external  
267 validation), the percentages were 92% and 88%, respectively.

268 It is important to highlight that these promising results were achieved by analyzing a set of samples  
269 that are representative of the large variety of parameters (olive cultivar, country of origin,  
270 commercial category) that can describe olive oil and affect its chemical characteristics.

271 It cannot be excluded *a priori* that other factors could also affect this discrimination, but the results  
272 clearly showed that the EU vs. extra-EU origin of the oils played a key role.

273 The results obtained herein sustain the use of multivariate chemometrics with untargeted detection  
274 of volatile compounds as a powerful tool to discriminate EVOOs and VOOs of different origin,  
275 leading to the possibility to verify compliance with the labelled geographical provenance (EU vs  
276 extra-EU). The identification of the specific country of origin, both in blended and not blended oils,  
277 may be of significant relevance, but to obtain a robust and satisfactory classification model a  
278 specific sampling, comprising a very high number of oils coming from each country, is still needed.

279 Other studies have already reported that the analysis of volatile compounds is suitable for tracing  
280 the geographical origin of VOOs. It is, therefore, of utmost importance to identify and quantify the  
281 main volatile compounds involved in the discrimination with other analytical techniques.

282 Nonetheless, the methodology proposed herein presents some advantages over other techniques  
283 generally applied for this analysis, as it is very rapid (only 200 s are needed for each  
284 chromatographic run) and easy to use since no sample treatment is required. In particular, for the  
285 effective application of the proposed analytical methodology, it would be worthwhile for each



286 laboratory to build its own model on a large dataset of samples of interest in order to achieve, year  
287 after year, a historical memory that is useful to control the conformity of oils coming from the same  
288 areas and sold by the same or different suppliers. Inter-laboratory tests should be also carried out to  
289 verify the comparability of the results among different laboratories, allowing the establishment of a  
290 shared data set to increase overall reliability, on which the predictive models should be based.

291

## 292 **Funding**

293 This work was supported by the Horizon 2020 European Research project OLEUM “*Advanced*  
294 *solutions for assuring the authenticity and quality of olive oil at a global scale*”, which received  
295 funding from the European Commission within the Horizon 2020 Programme (2014–2020), grant  
296 agreement no. 635690. The information expressed in this article reflects the authors’ views; the  
297 European Commission is not liable for the information contained herein.

298

## 299 **Acknowledgments**

300 The authors would like to thank the colleagues involved in the OLEUM project, from EUROFINs,  
301 IPTPO, ITERG, UZZK, and ZRS for collection and shipment of samples.

302

## 303 **References**

304

305 Angerosa, F., Bréas, O., Contento, S., Guillou, C., Reniero, F., & Sada, E. (1999). Application of  
306 stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *Journal*  
307 *of Agricultural and Food Chemistry*, 47, 1013-1017.

308

309 Bajoub, A., Medina-Rodríguez, S., Ajal, E. A., Cuadros-Rodríguez, L., Monasterio, R. P.,  
310 Vercammen, J., Fernández-Gutiérrez, A., & Carrasco-Pancorbo, A. (2018). A metabolic  
311 fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and

312 chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food*  
313 *Research International*, 106, 233-242.

314

315 Bajoub, A., Medina-Rodríguez, S., Hurtado-Fernández, E., Ajal E. A., Ouazzani, N., Fernández-  
316 Gutiérrez, & Carrasco-Pancorbo A. (2016). A first approach towards the development of  
317 geographical origin tracing models for North Moroccan olive oils based on triacylglycerols profiles.  
318 *European Journal of Lipid Science and Technology*, 118, 1223-1235.

319

320 Ben Mohamed, M., Rocchetti, G., Montesano, G., Ben Ali, S., Guasmi, F., Grati-Kamoun, G., &  
321 Lucini, L. (2018). Discrimination of Tunisian and Italian extra-virgin olive oils according to their  
322 phenolic and sterolic fingerprints. *Food Research International*, 106, 920-927.

323

324 Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford, UK: Oxford University  
325 Press.

326

327 Bontempo, L., Paolini, M., Franceschi, P., Ziller, L., García-González, D. L., & Camin, F. (2019).  
328 Characterisation and attempted differentiation of European and extra-European olive oils using  
329 stable isotope ratio analysis. *Food Chemistry*, 276, 782-789.

330

331 Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: taking the magic  
332 away. *Journal of Chemometrics*, 28, 213-225.

333

334 Chiocchini, F., Portarena, S., Ciolfi, M., Brugnoli, E., & Lauteri, M. (2016). Isoscapes of carbon  
335 and oxygen stable isotope compositions in tracing authenticity and geographical origin of Italian  
336 extra-virgin olive oils. *Food Chemistry*, 202, 291-301.

337

- 338 Commission implementing regulation (EU) No 29/2012 of 13 January 2012 on marketing standards  
339 for olive oil. *Official Journal of European Journal, L 12*, 14-21.
- 340
- 341 Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., Lacoste, F., Brereton, P., García-  
342 González, D. L., Moreda, W., & Gallina Toschi, T. (2019). Olive oil quality and authenticity: A  
343 review of current EU legislation, standards, relevant methods of analyses, their drawbacks and  
344 recommendations for the future. *Trends in Food Science & Technology*, in press.  
345 doi.org/10.1016/j.tifs.2019.02.025
- 346
- 347 Daszykowski, M., Walczak, B., Massart, D. L. (2002). Representative subset selection. *Analytica*  
348 *Chimica Acta*, 468, 91-103.
- 349
- 350 Dekhili, S., Sirieix, L., & Cohen, E. (2011). How consumers choose olive oil: the importance of  
351 origin cues. *Food Quality and Preference*, 22, 757-762.
- 352
- 353 Del Giudice, T., Cavallo, C., Caracciolo, F., & Cicia, G. (2015). What attributes of extra virgin  
354 olive oil are really important for consumers: a meta-analysis of consumers' stated preferences.  
355 *Agricultural and Food Economics*, 3, article number 20.
- 356
- 357 García-González D. L., Luna, G., Morales, M. T., & Aparicio, R. (2009). Stepwise geographical  
358 traceability of virgin olive oils by chemical profiles using artificial neural network models.  
359 *European Journal of Lipid Science and Technology*, 111, 1003-1013.
- 360
- 361 Gerhardt, N., Birkenmeier, M., Sanders, D., Rohn, S., & Weller P. (2017). Resolution-optimized  
362 headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil  
363 profiling. *Analytical and Bioanalytical Chemistry*, 409, 3933-3942.

364

365 Ghisoni, S., Lucini, L., Angilletta, F., Rocchetti, G., Farinelli, D., Tombesi, S., & Trevisan, M.  
366 (2019). Discrimination of extra-virgin-olive oils from different cultivars and geographical origins  
367 by untargeted metabolomics. *Food Research International*, *121*, 746-753.

368

369 Giacalone, R., Giuliano, S., Gulotta, E., Monfreda, M., & Presti, G. (2015). Origin assessment of  
370 EV olive oils by esterified sterols analysis. *Food Chemistry*, *188*, 279-285.

371

372 Kosma, I., Vatavali, K., Kontakos, S., Kontominas, M., Kiritsakis, A., & Badeka, A. (2017).  
373 Geographical differentiation of Greek extra virgin olive oil from late-harvested Koroneiki cultivar  
374 fruits. *Journal of the American Oil Chemists' Society*, *94*, 1373-1384.

375

376 Lukić, I., Carlin, S., Horvat, I., & Vrhovsek, U. (2019). Combined targeted and untargeted profiling  
377 of volatile aroma compounds with comprehensive two-dimensional gas chromatography for  
378 differentiation of virgin olive oils according to variety and geographical origin. *Food Chemistry*,  
379 *270*, 403-414.

380

381 Mallamace, D., Vasi, S., Corsaro, C., Naccari, C., Clodoveo, M. L., Dugo, G., & Cicero, N. (2017).  
382 Calorimetric analysis points out the physical-chemistry of organic olive oils and reveals the  
383 geographical origin. *Physica A*, *486*, 925-932.

384

385 Melucci, D., Bendini, A., Tesini, F., Barbieri, S., Zappi, A., Vichi, S., Conte, L., & Gallina Toschi,  
386 T. (2016). Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash  
387 gas chromatography electronic nose and chemometrics. *Food Chemistry*, *204*, 263-273.

388

389 Palagano, R., Bendini, A., Valli, E., Cevoli, C., Gallina Toschi, T. (2019). OLEUM Project.  
390 Analysis of volatile compounds of extra virgin and virgin olive oil with a different geographical  
391 origin by flash gas chromatography. <http://doi.org/10.6092/unibo/amsacta/6243> [Dataset]

392

393 Peršurić, Ž., Saftić, L., Mašek, T., & Kraljević Pavelić, S. (2018). Comparison of triacylglycerol  
394 analysis by MALDI-TOF/MS, fatty acid analysis by GC-MS and non-selective analysis by NIRS in  
395 combination with chemometrics for determination of extra virgin olive oil geographical origin. A  
396 case study. *LWT - Food Science and Technology*, 95, 326-332.

397

398 Quintanilla-Casas, B., Bertin, S., Leik, K., Bustamante, J., Guardiola, F., Valli, E., Bendini, A.,  
399 Gallina Toschi, T., Tres, A., & Vichi, S. (2020). Profiling versus fingerprinting analysis of  
400 sesquiterpene hydrocarbons for the geographical authentication of extra virgin olive oils. *Food*  
401 *Chemistry*, 307, Article number 125556.

402

403 Sayago, A., González-Domínguez, R., Beltrán, R., & Fernández-Recamales, A. (2018).  
404 Combination of complementary data mining methods for geographical characterization of extra  
405 virgin olive oils based on mineral composition. *Food Chemistry*, 261, 42-50.

406

407 Tomasi, G., Van Den Berg, F., & Andersson, C. (2011). Correlation optimized warping and  
408 dynamic time warping as preprocessing methods for chromatographic data. *Journal of*  
409 *Chemometrics*, 18, 231-241.

410

411 Valli, E., Bendini, A., Berardinelli, A., Ragni, L., Riccò, B., Grossi, M., & Gallina Toschi, T.  
412 (2016). Rapid and innovative instrumental approaches for quality and authenticity of olive oils.  
413 *European Journal of Lipid Science and Technology*, 118, 1601-1619.

414

415 Vichi, S., Pizzale, L., & Conte, L. (2007). Stereospecific distribution of fatty acids in  
416 triacylglycerols of olive oils. *European Journal of Lipid Science and Technology*, 109, 72-78.

Journal Pre-proof

**Table 1.** Number of samples for each class of origin considered and geographical origin. EU: oils from EU state members. Extra-EU: oils from countries outside the European Union. Blends: oils obtained mixing EU oils or EU and extra-EU oils.

<b>Origin class</b>	<b>N</b>	<b>Country of origin</b>
EU	116	29 Spain, 25 Italy, 22 Croatia, 16 Greece, 12 Portugal, 12 Slovenia
Extra-EU	70	42 Morocco, 21 Turkey, 6 Tunisia, 1 Chile
Blends	24	12 EU blends, 12 EU/extra-EU blends

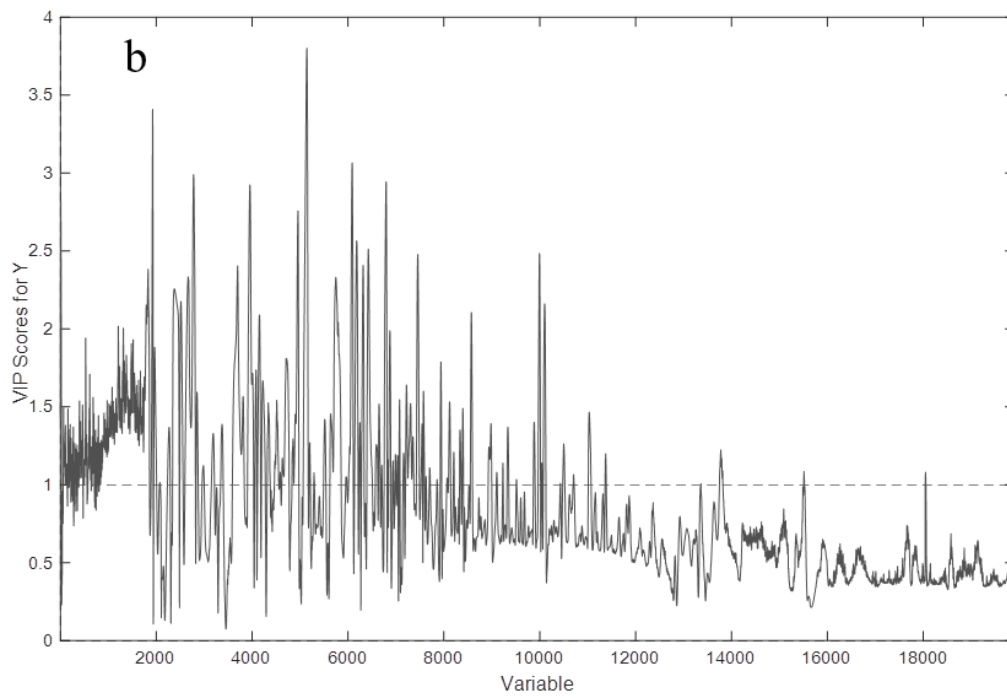
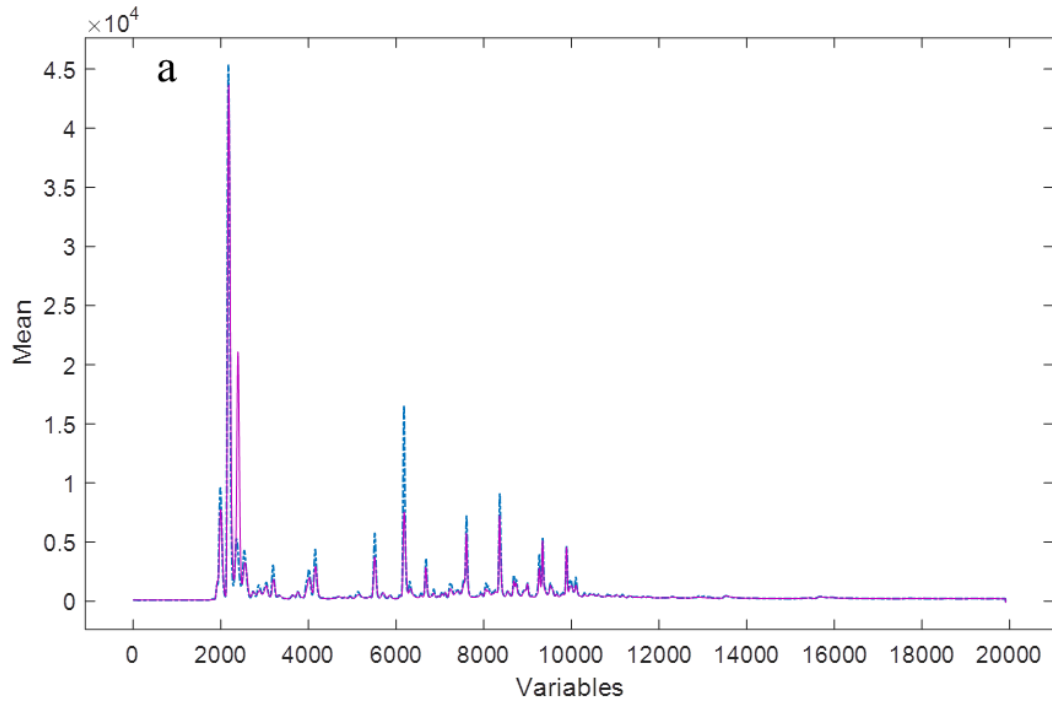
**Table 2.** Percentages and number (in parentheses) of correctly classified samples for each category using the PLS-DA model. EU: oils from a single state member of European Union and oils obtained by mixing EU oils. Extra-EU: oils from a single country outside the European Union and oils obtained by mixing EU and extra-EU oils.

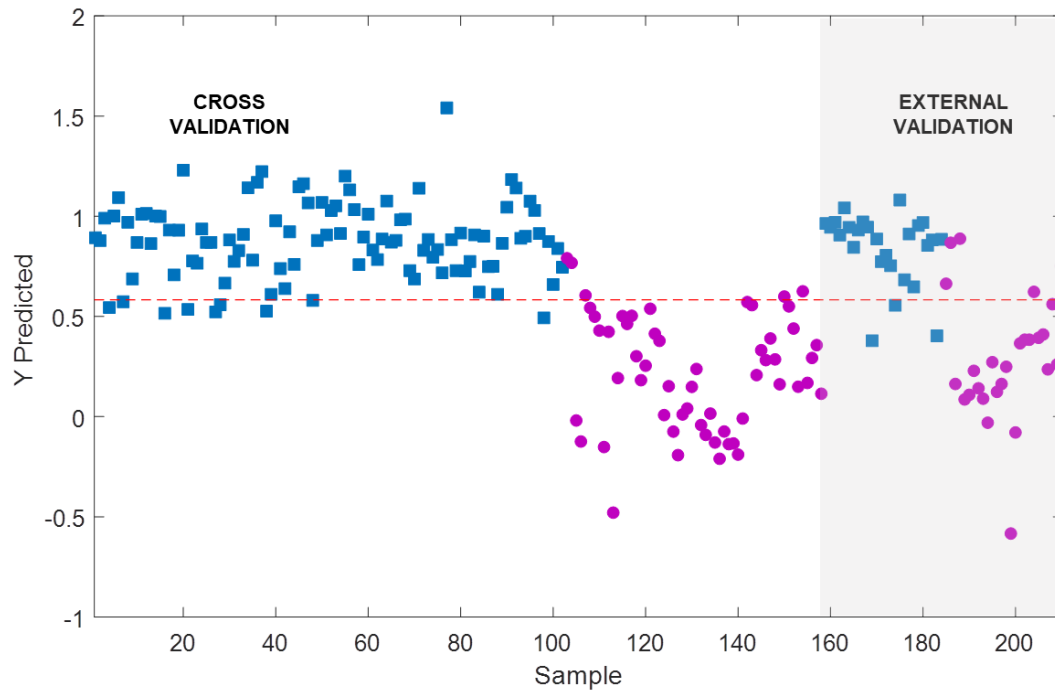
<b>Category</b>	<b>Cross validation</b>	<b>External validation</b>
EU	91.2% (93/102)	88.5% (23/26)
Extra-EU	91.1% (51/56)	80.8% (21/26)

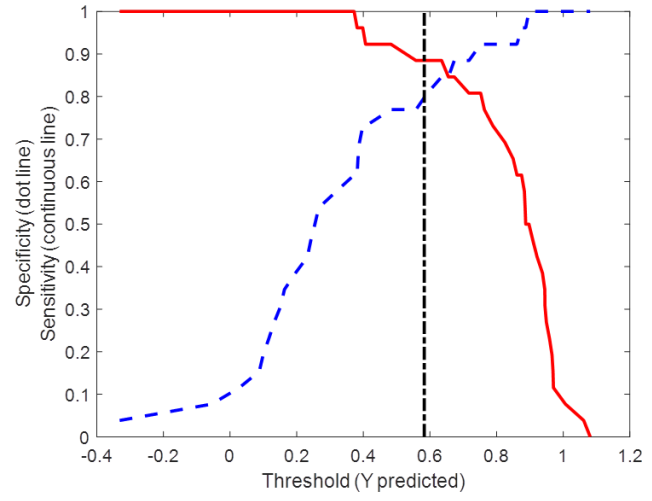
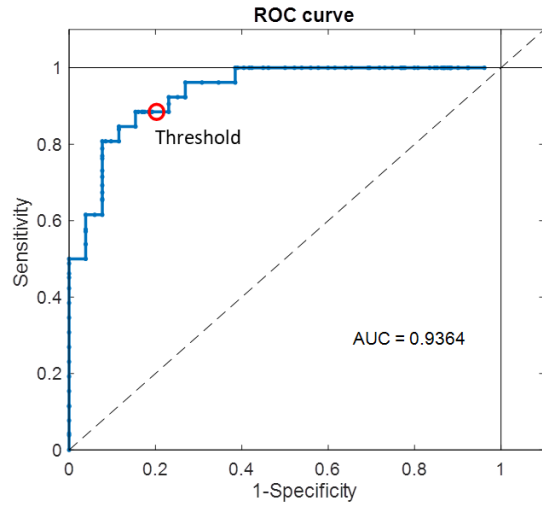


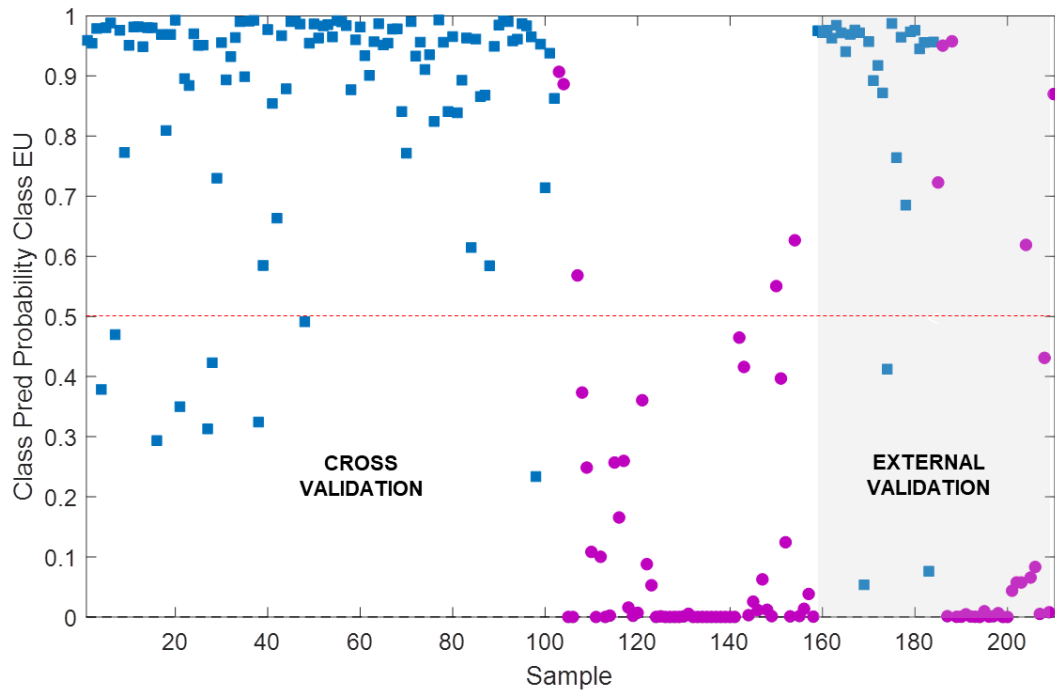
**Table 3.** Percentages of samples correctly classified for each category using the ANN model. EU: oils from a single state member of the European Union and oils obtained by mixing EU oils. Extra-EU: oils from a single country outside the European Union and oils obtained by mixing EU and extra-EU oils.

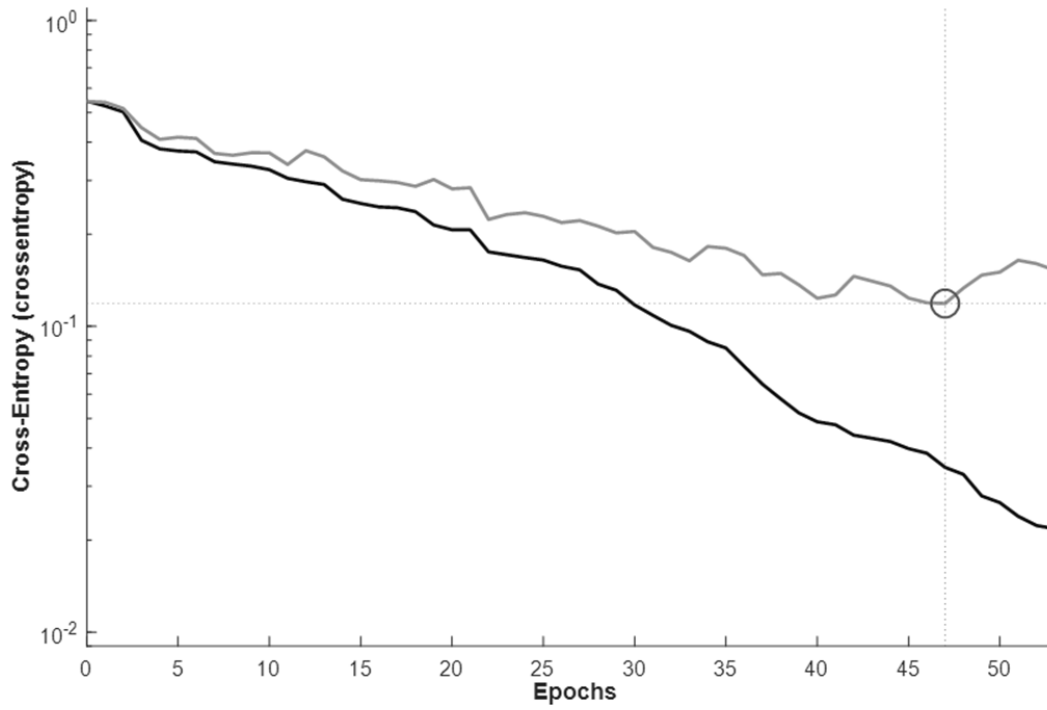
<b>Category</b>	<b>Training (%)</b>	<b>Verification (%)</b>	<b>External validation (%)</b>
EU	97.5 (80/82)	95.0 (19/20)	92.3 (24/26)
Extra-EU	93.2 (41/44)	91.7 (11/12)	88.4 (23/26)

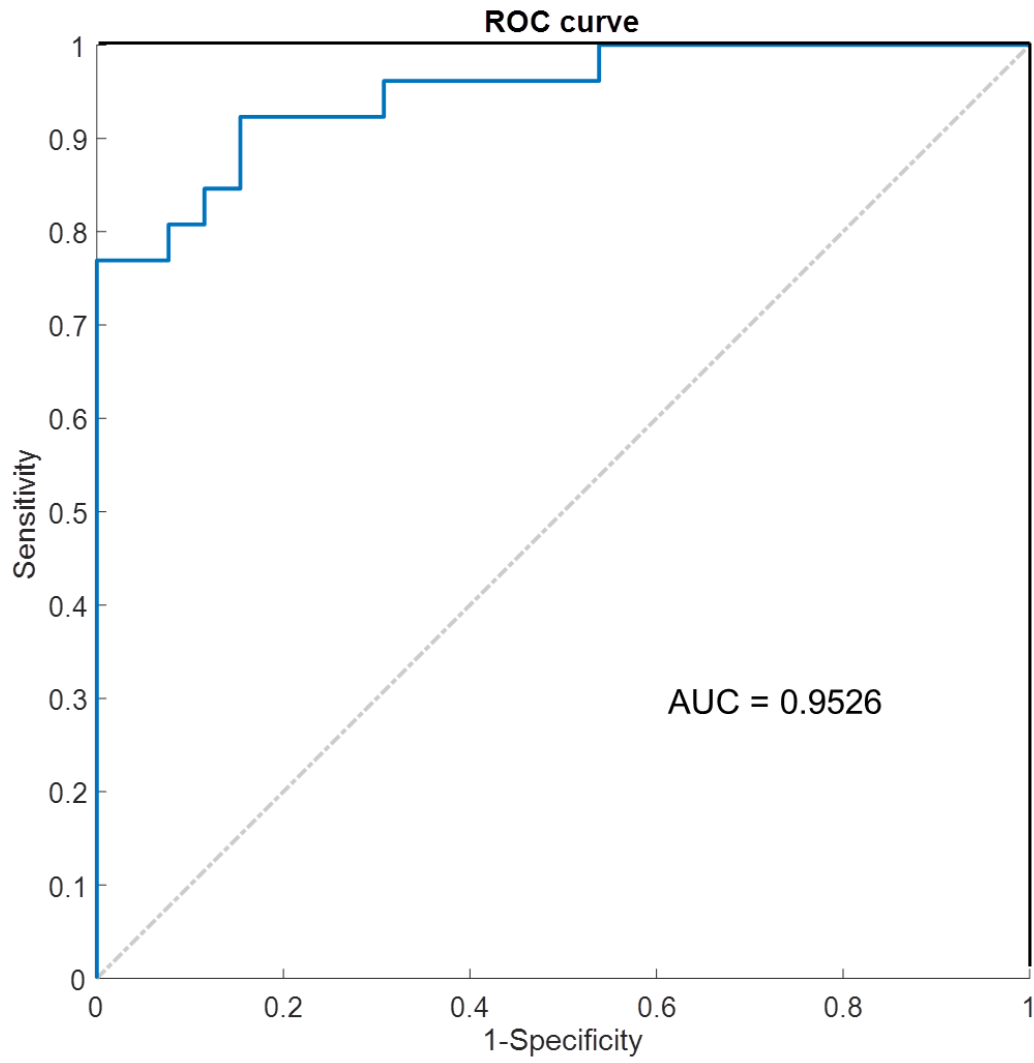












## Highlights

- FGC was applied to discriminate oils with different geographical origin
- Chemometric analysis with untargeted approach was applied
- PLS-DA and ANN classification models were built and validated
- Satisfactory percentages of samples correctly classified were obtained



Conflict of Interest and Authorship Conformation Form

Please check the following as appropriate:

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript