Artificial Intelligence Approaches for Fast and Portable Traceability Assessment of EVOO

Luciano Ortenzi, Simone Figorilli, Simona Violino,
Federico Pallottino and Corrado Costa

July 19, 2023

Artificial Intelligence approaches for fast and portable traceability assessment of EVOO

***Abstract*— Extra virgin olive oil (EVOO) represents one of the first-choice products made in Italy for its high quality and use in the Mediterranean diet. The aim of this study was to evaluate the effectiveness of a portable VIS–NIR open-source spectroscopic system coupled with an artificial intelligence model for the rapid determination of EVOOs traceability. Reported results for EVOO traceability, with respect to different degrees of aggregation (EU and extra-EU, Italian and foreign and Italian areas of membership), show excellent performances of artificial intelligence models and indicate a valid rapid and low-cost method of analysis for combating EVOO counterfeiting.**

Keywords—IoT, Made in Italy, ANN, Machine learning, Agriculture, Food

I. INTRODUCTION

Extra virgin olive oil (EVOO) is one of the most important ingredients in the Mediterranean diet, being used by most countries in the Mediterranean basin due to its excellent qualities and sensory properties attributable to the fruits of the olive tree (*Olea europaea* L.) [1]. The price of EVOO is on average 4-5 times higher than that of other vegetable oils. This is due to the higher production costs and its higher nutritional and organoleptic properties. In principle, therefore, the higher cost should help ensure high quality standards [2]. However, within a globalized market, this product is subject to continuous counterfeiting; therefore, it is necessary to raise consumer awareness of the peculiarities and qualities that distinguish “Made in Italy” from other products. Useful tools to combat counterfeiting are those that ensure traceability [3], or the “possibility of reconstructing and following the path of a food at all stages of production, processing and distribution” [4]. It is important to use accurate analytical methods that can guarantee the authenticity and traceability of foods, especially for those presenting high quality and economic value such as extra virgin olive oil. Several analytical techniques have been developed over the past decade to aid in the identification of olive oil, and about 200 compounds, out of hundreds, have proven useful as compositional markers for the purpose of EVOO traceability. The compositional markers include both major and minor components. State-of-the-art approaches to EVOO traceability for geographic origin assessment include determination of major components (e.g., triglycerides and fatty acids), stable isotope ratios (e.g., 13C/12C in combination with 18O/16O), and multi-element characterization through the application of various multivariate statistical techniques and artificial intelligence (AI). Those commonly used for data analysis are cluster analysis, multidimensional scaling, artificial neural networks (ANNs), and partial least squares discriminant analysis (PLSDA) [5]. Among the various techniques to identify the authenticity and traceability of EVOO, there are non-destructive solutions such as infrared spectroscopy (IRS). The combination with machine learning techniques has been used satisfactorily for authentication and screening of olive oil [5,6]. As reported by Sayago et al. [7] machine learning algorithms have been used to build classification and prediction models to determine variety and geographical origin of olive oils collected from different locations in the province of Huelva (Spain). In the work of Funes et al. [8] neural networks were used to analyze the NIR spectra of the oil must as it exited the horizontal decantation centrifuge and the flows and temperatures of the oil and addition water in order to control the quality of the EVOO. In that work all designed networks showed very good performances. Another work showed the effectiveness of artificial neural networks to track the authenticity (according to variety, origin and cut) of Taggiasca Ligure, a renowned Italian EVOO. Neural networks enabled the identification of specific markers of authenticity namely cholesterol derivatives and phenols [9]. The purpose of this work was to assess the actual geographic origin of EVOOs labeled on the market as Italian and to test the potential efficiency of an open source AI-powered VIS-NIR device for traceability purposes. Indeed, the device could produce results for olive oil authentication (by variety and origin) and fraud detection in a fraction of the time and potentially on a much larger number of samples than conventional analytical methods. In detail, the study pursued the objective of analysing the geographical origin of 203 EVOO samples produced between 2018 and 2020 with respect to three different degrees of aggregation: 1) EU vs non-EU 2) Italian non vs non Italian 3) Italian areal distribution. This was done by collecting the optical spectra of EVOOs and analysing them through an ANN. To investigate a broader range of optical frequencies, two portable spectrometers were used.

II. MATERIALS AND METHODS

A. Evoo samples

To determine the EVOO traceability, in this study a total of 203 samples of worldwide EVOOs were collected and the Italian ones came from different part of Italy (North, Centre, South, Islands). The EVOO samples were analised with two different VIS-NIR spectrometers (see description below) and the collected spectra were grouped according to different degree of aggregation: EU and EXTRA-EU, Italian and foreign and according to different Italian areal distribution.

B. IoT Spectrometers

Oil samples were analyzed with two VIS–NIR spectrometers: Lumini C (My spectral LTD., Cambridge, MA, USA) for tasks 1) and 3) and the SCIO^{TM} (TM in apice) by consumer physics, for task 2). The SCIO™ spectrometer: a low-cost ready-to-use, pocket-sized sensor that uses a short wavelength NIR range (Figure 1).

Fig. 1. VIS–NIR spectrophotometer SCIO^{TM}

The collected reflectance signal covers a range of 740–1070 nm. For each sample, 15 spectral replicates were acquired and then averaged. The device uses Bluetooth wireless technology to communicate with a smartphone and the “SCIO Lab” app. The spectral data were sent to cloud server by “SCIO Lab” app and can be exported via SCIO web portal experiment in a CSV format and transferred to an excel sheet for analysis.

The ultra-compact VIS-NIR spectrophotometer shown in Figure 2 is the Lumini C (Myspectral Ltd., Cambridge, MA, USA), able to measure spectral reflectance or absorbance. The device is small, light, low-cost, and open source. The spectral range covered is 340–890 nm with an optical resolution equal to 8 nm and wavelength accuracy equal to 0.5 nm. For appropriate acquisition of the spectral signature, in relation to the sample reflectance characteristics, the acquisition can be set at different integration times. A specific Android app has been developed for the multi-acquiring of spectral data sent by the Lumini C device and storaged by a local database [5].

Fig. 2. VIS-NIR ultra-compact spectrophotometer Lumini C Myspectral.

C. Machine learning models and feature importance

In order to assign the collected optical spectra to the correct categories considered, three machine learning models were built in order to 1) discriminate between EU and non-EU olive oil, 2) recognize Italian and Foreign EVOOs 3) assign the correct Italian areal range. In all three cases a single hidden layer artificial neural network with a given number of nodes (N) was trained with a given number of epochs (E) and applied. For the three cases the training was done using the scaled conjugate gradient back propagation algorithm [10] as implemented in the deep learning MATLAB (The MathWorks Inc. MA USA) toolbox, with a threshold on the gradient equals to 10^{-10} and using the mean square error (MSE) as performance function. For all three cases the dataset of 203 samples was random partitioned using an 80 to 20 ratio between training and test sets.

Training hyperparameters like N and E are reported in table 1 for the three cases of interest. To interpret the results obtained the feature importance was also calculated as described in Navarro *et al.* [11].

TABLE 1. MAIN FEATURES OF THE MODELS USED.

Parameters	***EU vs. EXTRA-EU***	***Italian vs. Foreign***	***Areal distribution***
#input variables	288	331	288
% training set	80	80	80
% test set	20	20	20
N. of hidden layers	10	50	70
Training Function:	Scaled Conjugate gradient	Scaled Conjugate gradient	Scaled Conjugate gradient
Performance function:	Mean squared error	Mean squared error	Mean squared error
Epochs	5000	7000	5000

III. RESULTS

From the results obtained from Italian and Foreign EVOO, through ANN analysis, only 6 samples out of 203 were misclassified (97.3% correctly classified) (Figure 3). The spectral range of greatest significance is between 740-900 nm (Figure 4). The samples that were misclassified belonged to the three collection campaigns and came from large retailers.

Fig. 3. Confusion Matrix for training and test set.

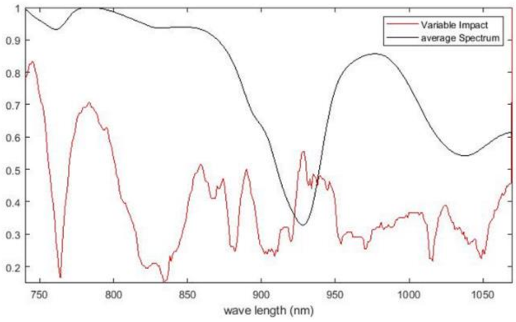


Fig. 4. “Variable impact” on the model (red line), average NIR spectrum of the analyzed samples (black line). For better comparison all quantities were normalized to maximum.

In the case study of the traceability of EU and non-EU EVOO samples, the best results were obtained through the VIS-NIR Lumini C device. In fact, ANN analysis showed that all samples were correctly classified (100%), (Figure 5). The spectral range of greatest significance is between 340-500nm (Figure 6).

Training Confusion Matrix

	1	2	
1	16 9.9%	0 0.0%	100% 0.0%
2	0 0.0%	146 90.1%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%

Validation Confusion Matrix

	1	2	
1	0 NaN%	0 NaN%	NaN% NaN%
2	0 NaN%	0 NaN%	NaN% NaN%
	NaN% NaN%	NaN% NaN%	NaN% NaN%

Test Confusion Matrix

	1	2	
1	1 2.4%	0 0.0%	100% 0.0%
2	0 0.0%	40 97.6%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%

All Confusion Matrix

	1	2	
1	17 8.4%	0 0.0%	100% 0.0%
2	0 0.0%	186 91.6%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%

Fig. 5. Confusion Matrix for the training set and for the test.

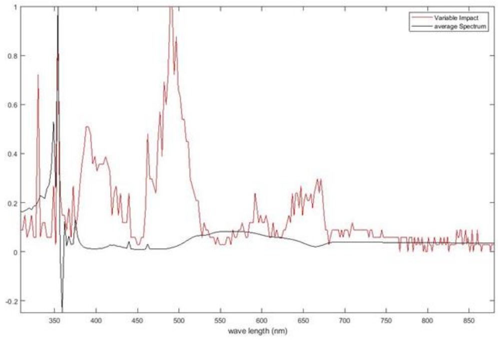
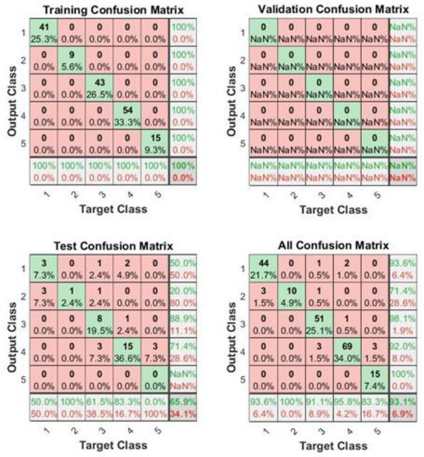


Fig. 6. “Variable impact” on the model (red line), average NIR spectrum of the analyzed samples (black line). For better comparison all quantities were normalized to the maximum.

Regarding areal ranges, from the results obtained, through ANN analysis on spectral data obtained through the VIS-NIR Lumini C device, only 14 samples were misclassified (93.1% correctly classified), (Figure 7). The spectral range of greatest importance is between 330-500 nm (Figure 8). In this study, 14 samples were misclassified; however, this represents a good result given the complexity of the areas (4 different areas according to the UNAPROL Handbook – Control Systems and Compliance with Authenticity and Quality Standards for Olive Oil and Table Olives). In addition, by increasing the number of samples, the model can be trained better and have fewer classification errors.



Training Confusion Matrix

	1	2	3	4	5	
1	41 25.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	9 5.6%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	43 26.5%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	54 33.3%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	15 9.3%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Validation Confusion Matrix

	1	2	3	4	5	
1	0 NaN%	0 NaN%	0 NaN%	0 NaN%	0 NaN%	NaN% NaN%
2	0 NaN%	0 NaN%	0 NaN%	0 NaN%	0 NaN%	NaN% NaN%
3	0 NaN%	0 NaN%	0 NaN%	0 NaN%	0 NaN%	NaN% NaN%
4	0 NaN%	0 NaN%	0 NaN%	0 NaN%	0 NaN%	NaN% NaN%
5	0 NaN%	0 NaN%	0 NaN%	0 NaN%	0 NaN%	NaN% NaN%
	NaN% NaN%	NaN% NaN%	NaN% NaN%	NaN% NaN%	NaN% NaN%	NaN% NaN%

Test Confusion Matrix

	1	2	3	4	5	
1	3 7.3%	0 0.0%	1 2.4%	2 4.9%	0 0.0%	50.0% 50.0%
2	3 7.3%	1 2.4%	1 2.4%	0 0.0%	0 0.0%	20.0% 80.0%
3	0 0.0%	0 0.0%	8 19.5%	1 2.4%	0 0.0%	88.9% 11.1%
4	0 0.0%	0 0.0%	3 7.3%	15 36.6%	3 7.3%	71.4% 28.6%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	50.0% 50.0%	100% 0.0%	61.5% 38.5%	83.3% 16.7%	0.0% 100%	65.9% 34.1%

All Confusion Matrix

	1	2	3	4	5	
1	44 21.7%	0 0.0%	1 0.5%	2 1.0%	0 0.0%	93.6% 6.4%
2	3 1.5%	10 4.9%	1 0.5%	0 0.0%	0 0.0%	71.4% 28.6%
3	0 0.0%	0 0.0%	51 25.1%	1 0.5%	0 0.0%	98.1% 1.9%
4	0 0.0%	0 0.0%	3 1.5%	69 34.0%	3 1.5%	92.0% 8.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	15 7.4%	100% 0.0%
	93.6% 6.4%	100% 0.0%	91.1% 8.9%	95.8% 4.2%	83.3% 16.7%	93.1% 6.9%

Fig. 7. Confusion matrix for the training set and for the test.

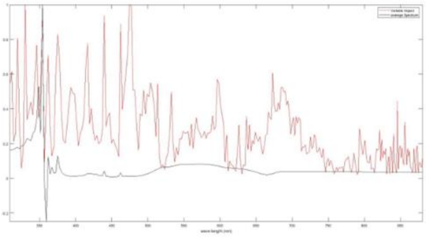


Fig. 8. “Variable impact” on the model (red line), average NIR spectrum of the analyzed samples (black line). For better comparison all quantities were normalized to the maximum.

IV. CONCLUSIONS

The work enabled the determination of EVOO traceability through the combined use of spectrophotometry and AI. In detail, two open-source and inexpensive tools were used that, compared to other techniques such as chromatography or chemical analysis, allow results to be obtained in very little time and with a minimal economic budget. Thanks to the AI models on the sample matrices, excellent performances were observed in all three datasets used. The present work represents a paradigmatic example of an IoT implementation including a combination of low-cost sensors, cloud computing and intelligent signal elaboration. It is worth noticing that VIS-NIR spectrometers are widely cheaper than the far infrared (FIR) ones. Moreover, even data collection is easier to perform, since samples do not need to be treated in special ways. However, the VIS-NIR signal interpretation is far more difficult due to the presence of noise coming from several sources included the presence of overtones in the optical spectrum. As a result, the machine learning part is crucial in making the workflow efficient, fast and easy-to-use. The present protocol prototype allows for an autonomous fast traceability-check by the single EVOO consumer which is free to choose, and eventually pay a higher prize for an authentic product. As a result, the sellers become less prone to frauds and arbitrary increase of prizes. AI-powered spectrometers can really help to ensure the authenticity and quality of EVOOs, preventing consumers from the increasing counterfeits in the EVOO market.

Acknowledgment

The authors would like to acknowledge all the farms which sent the EVOO to be analysed. This study was funded by the Italian Ministry of Agriculture (MiPAAF) in the project INFOLIVA (D.M. n. 12479).

REFERENCES

[1] M. Benito, R. Oria, A.C. Sánchez-Gimeno, “Characterization of the olive oil from three potentially interesting varieties from Aragon (Spain)”, FOOD SCIENCE AND TECHNOLOGY INTERNATIONAL, 16: 523–530, 2010.

[2] S. Violino, F. Pallottino, G. Sperandio, S. Figorilli, F. Antonucci, V. Ioannoni, D. Fappiano, C. Costa, “Are the innovative electronic labels for extra virgin olive oil sustainable, traceable, and accepted by consumers?” FOODS, 8:529, 2019.

[3] S. Violino, F. Antonucci, F. Pallottino, C. Cecchini, S. Figorilli, C. Costa, “Food traceability: A term map analysis basic review”, EUROPEAN FOOD RESEARCH AND TECHNOLOGY, 1–11, 2019.

[4] European Commission. Regulation (EC) No 178/2002 of the European parliament and of the council. Laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Union, 31:1–24, 2002.

[5] S. Violino, L. Ortenzi, F. Antonucci, F. Pallottino, C. Benincasa, S. Figorilli, C. Costa,“An artificial intelligence approach for Italian EVOO origin traceability through an open source IoT spectrometer”, FOODS, 9(6): 834, 2020.

[6] C. Scotigno and G. Festa, “FTIR coupled with machine learning to unveil spectroscopicbenchmarks in the Italian EVOO”, International Journal of Food Science and Technology, 57, 4156–4162 (2022).

[7] A. Sayago, R. González-Domínguez, J. Urbano, A. Fernández-Recamales, “ Combination of vintage and new-fashioned analytical approaches for varietal and geographical traceability of olive oils” FOOD SCIENCE AND TECHNOLOGY, 111: 99-104, 2019.

[8] E. Funes, Y. Allouche, G. Beltrán, M.P. Aguliera, A. Jiménez, “Predictive ANN models for the optimization of extra virgin olive oil clarification by means of vertical centrifugation”, JOURNAL OF FOOD PROCESS ENGINEERING, 41(1): e12593, 2019.

[9] B. Senizza, P. Ganugi, M. Trevisan, L. Lucini, “Combining untargeted profiling of phenolics and sterols, supervised multivariate class modelling and artificial neural networks for the origin and authenticity of extra-virgin olive oil: A case study on Taggiasca Ligure”, FOOD CHEMISTRY, 404: 134543, 2023.

[10] M.F. Møller, “Efficient training of feed-forward neural networks” DAIMI REPORT SERIES, 464, 1993.

[11] A. Navarro, N. Nicastro, C. Costa, A. Pentangelo, M. Cardarelli, L. Ortenzi , F. Pallottino , T. Cardi and C. Pane,“ Sorting biotic and abiotic stresses on wild rocket by leaf-image hyperspectral data mining with an artifcial intelligence model”, Plant Methods (2022) 18:45, 2020.