Interreg
Greece-Italy
AUTHENTIC-OLIVE-NET
European Regional Development Fund
EUROPEAN UNION

ΕΛΓΟ-ΔΗΜΗΤΡΑ
ΕΛΛΗΝΙΚΟΣ ΓΕΩΡΓΙΚΟΣ
ΟΡΓΑΝΙΣΜΟΣ - ΔΗΜΗΤΡΑ

Annex II

Template for Studies and Researches

“Certification of Authenticity and Development of a Promotion Network olive products in the across border GREECE – ITALY area”

“AUTHENTIC-OLIVE-NET”

Development of a machine learning model for olive cultivar classification

This study was prepared by:

Dr.Ioannis Manousopoulos, Dr. Vasiliki Skiada and Panagiotis Katsaris

This document has been produced with the financial assistance of the European Union. It reflects the author’s views and the Managing Authority shall not be held liable for any use that may be made of the information contained therein.

The partnership of "AUTHENTIC-OLIVE-NET"

BENEFICIARY No	BENEFICIARY INSTITUTION	Role	Country
LB1	PREVEZA CHAMBER	Lead Beneficiary	Greece
PB2	REGION OF WESTERN GREECE	Partner	Greece
PB3	HELLENIC AGRICULTURAL ORGANIZATION - DEMETER	Partner	Greece
PB4	ASSOPROLI BARI AGRICULTURAL COOPERATIVE SOCIETY	Partner	Italy
PB5	CHAMBER OF COMMERCE OF FOGGIA	Partner	Italy

History Changes

Version Number	Date of Issue	Author(s)
1	31-03-2022	Dr. Y. Manousopoulos, Dr. V. Skiada and P. Katsaris

Table of Contents

1. EXECUTIVE SUMMARY

Olive oil samples of five monovarietal cultivars collected from several locations in the northwestern part of Greece (cv. Koroneiki, cv. Lianolia) and the southern peninsular of Italy (cv. Coratina, cv. Peranzane, cv. Favoloza) for three consecutive cultivation years were analyzed for their chemical composition in order to identify patterns for olive oils botanical classification and investigate the potential of developing a classification model capable of olive cultivar prediction.

Box plot analysis revealed patterns in the examined olive oil chemical compounds (sterol and fatty acid-28 variables) some of which remained constant between years (Figures 1,2,3,4). For the identification of potential trends between the examined chemical compounds and olive varieties, PCA was performed**.** The first and second components explained about 40% of the variance revealing trends in variables. The Greek varieties were totally separable to each other or from the Italian varieties (Figures 5, 6, 7). In contrast the Italian varieties showed a permanent overlapping to each other and an overlapping pattern with the Greek variety "Koroneiki" when all data from the three cultivation years were used.

Further statistical analysis (regression and non-parametric tests) were used to explore the potential effect of "Variety" and "Sampling period" on each of the 28 olive oil chemical compounds. The effects of varieties were also examined separately for the Greek and Italian varieties (Supplements material-1).Concerning stability of the examined chemical compounds across years, the Greek varieties had less significant differences in multiple comparisons compared to the Italian varieties (Supplements material-1). As expected by the PCA results, differences among the Italian varieties, which tended to overlap, were less significant than differences between the Greek varieties, which were almost totally separable (Supplements material-1).

The final step was the development of a classification algorithms for botanical classification The xgboost algorithm gave the best results and was used for model development. The developed model, had over 99% accuracy in predicting most olive varieties. Best results were obtained with the Greek varieties which were separated with 99.9 % accuracy. The Italian varieties were also classified with high accuracy with the exception of the Favolosa variety, in which accuracy dropped to 80% most likely due to small sample size.

Further in-depth research enriched with a higher number of samples and for more cultivation periods will improve the Machine learning model by stabilizing the variance of the examined olive oil compounds. As this work belongs to the first systematic attempt focusing on botanical classification (Tahir, H.E.,2022) with the use of machine learning model (xgboost algorithm), further investigation is under way which will possible increase flexibility by selecting the least required number of the most informative olive oil compounds on the examined dataset

In conclusion, this study can contribute in the future to the establishment of a continuously enriched "Authentic Olive Network" model of indigenous, local, and unexploited monovarietal olive oils. Finally another promising extension of this work, with expected applications in analogous fields of the food industry (e.g. cheese) could be the training of artificial intelligence models for the detection of olive oil adulteration.

.

2. SAMPLING

2.1 Sample collection and dataset preparation

Olive oil samples from 5 different olive varieties, cultivated in the northwestern part of Greece (cv. Koroneiki, cv. Lianolia Kerkyras) and the southern peninsular of Italy (cv. Coratina, cv. Peranzana and cv. Favoloza) were collected in three consecutive harvest periods. A total of 224 and 161 olive oil samples were collected during the first and the second harvest period respectively in order to create the dataset used for model training for olive cultivar prediction. In the extension (third) period 20 more Greek olive oil samples were collected. The total number of samples collected during project implementation are shown in Table 1, and their distribution across periods is shown in Table 2.

Table 1 Number of olive oil samples per variety

Variety	Samples
Koroneiki	77
Lianolia	90
Coratina	100
Favolosa	58
Peranzana	80
	405

Table 2 Number of olive oil samples collected in three harvest periods from Greece and Italy

a/a	Period	Country	Variety	N
1	1st	Greece	Koroneiki	44
2	1st	Greece	Lianolia	60
3	1st	Italy	Coratina	61
4	1st	Italy	Favolosa	19
5	1st	Italy	Peranzana	40
6	2nd	Greece	Koroneiki	21
7	2nd	Greece	Lianolia	22
8	2nd	Italy	Coratina	39
9	2nd	Italy	Favolosa	39
10	2nd	Italy	Peranzana	40
11	3rd	Greece	Koroneiki	12
12	3rd	Greece	Lianolia	8
			Total	**405**

Table 3 Chemical compounds of olive oil used in the study

Faty acids	Sterols
C14:0	24_meth_cholesterol
C16:0	b_sitosterol
C16:1	campestanol
C17:0	campesterol
C17:1	chlerosterol
C18:0	Cholestanol
C18:1	Δ5,24_stigm/dienol
C18:2	Δ5-avenasterol
C18:3	Δ7-avenasterol
C20:0	Δ7-stigmastenol
C20:1	sitostanol
C22:0	stigmasterol
C24:0	Total.β_sitosterol
	Total.erythrodiol
	Total sterols

Samples were collected directly from the cooperative local olive mills, following the same olive oil extraction procedure as arranged in a fixed instruction protocol. The obtained olive oil samples were labelled accordingly, and transferred directly at two accredited laboratories in Greece and Italy until chemical analysis were performed.

A total of 34 olive oil chemical characteristics were analyzed, including acidity, peroxide value, K232, K268, sterolic content and fatty acid composition. All chemical analysis were performed following the official analytical methodology (Commisiton Regulation EEC/2568/91).The main qualitative indices of acidity, peroxide value and spectroscopic measurents were performed in order to classify olive oil samples according to their category. All samples belonged to the highest category of “Extra Virgin Olive Oil”. The chemical parameters of fatty acid and sterolic profile (28 variables) were used as the exploratory variables of the dataset (Conte et.al 2020, Lozano-Castellón et.al.2022). The dataset was curated for missing values and outliers by coding functions of the “tidyverse” library of the R statistical language. There were 20 missing values in the “C24:0” variable and one missing value in the “Δ5-avenasterol” variable, which were imputated by the predictive mean matching algorithm of the mice “R” library. Five variables were excluded due to curation issues. The final set of exploratory variables comprised two main categories consisting of 15 sterols (individual sterols, total sterols and triterpen dialchohols) and 13 fatty acid compounds (Table 3). The olive dataset was used for explorative statistical analysis and was the basis for machine learning models development.

3. Statistical Analysis

Statistical analysis involved exploratory methods (EDA), principal components and regression analysis and was performed by using or programming functions of the following libraries of the “R” statistical language (R Core Team 2018): “tidyR” (Hadley Wickham 2021); “dplyr” (Hadley Wickham et al. 2021); ggplot2 (Hadley Wickham 2016); “emmeans” (Russell V. Lenth 2021); “performance” (Daniel Ludecke et al. 2021); “mice” (Stef van Buuren and Karin Groothuis-Oudshoorn 2011);“caret” (Max Kuhn 2020); “openxlsx” (Philipp Schauberger and Alexander Walker 2021); “factoextra” (Alboukadel Kassambara and Fabian Mundt 2020); “FactoMiner” (Sébastien Lê et al. 2008); “xgboost” (Tianqi Chen et al. 2021).

3.1 Box Plots

The sterolic and fatty acid profile (28 olive oil compounds) of the examined olive varieties were compared for each harvest period using boxplots (Figures 1, 2). Patterns between periods showed considerable similarity. Most olive oil compounds of the Italian olive varieties had little within period differences. In contrast, the two Greek varieties differed considerably in most of their chemical compounds within and between years. In both cases there was extensive variation in most of the examined olive oil compounds between years, most likely due to environmental and agronomic factors.

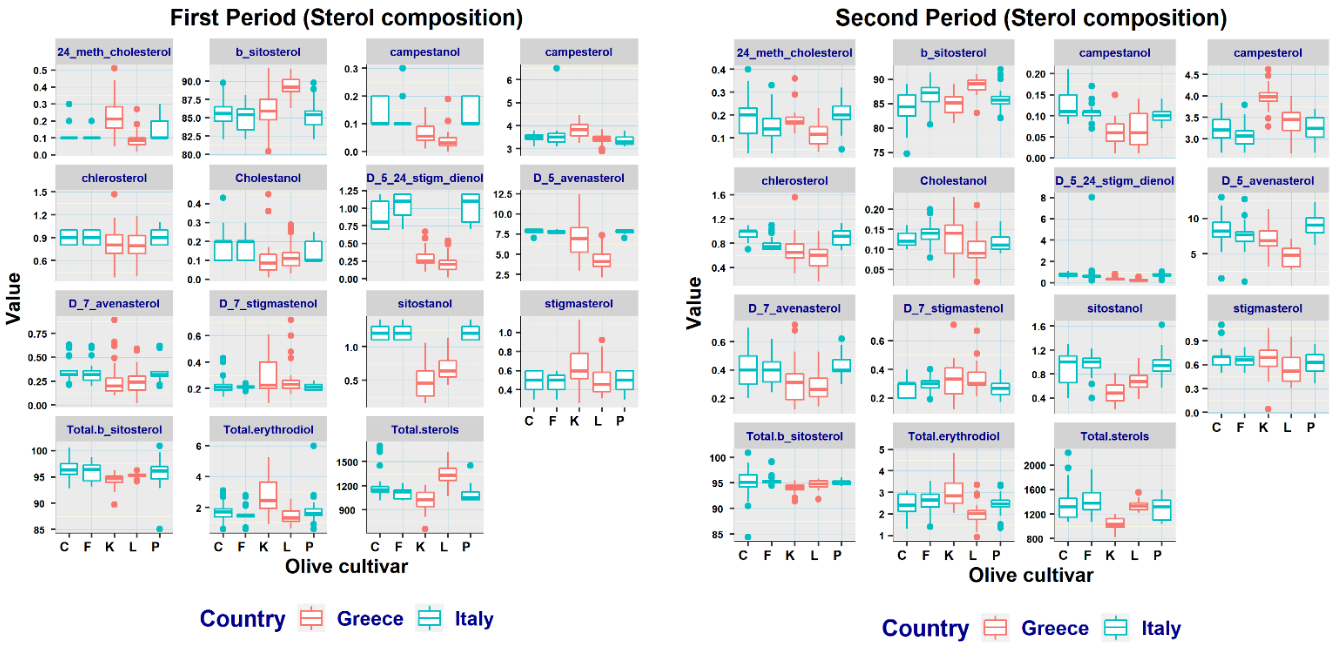

Figure 1. Box plot description of the sterol variables of the first and second sampling periods. C, F, K, L, P stand for Coratina, Favolosa, Koroneiki, Lianolia and Peranzana, respectively. Individual sterols are expressed in (%) and total sterols in mg/kg.

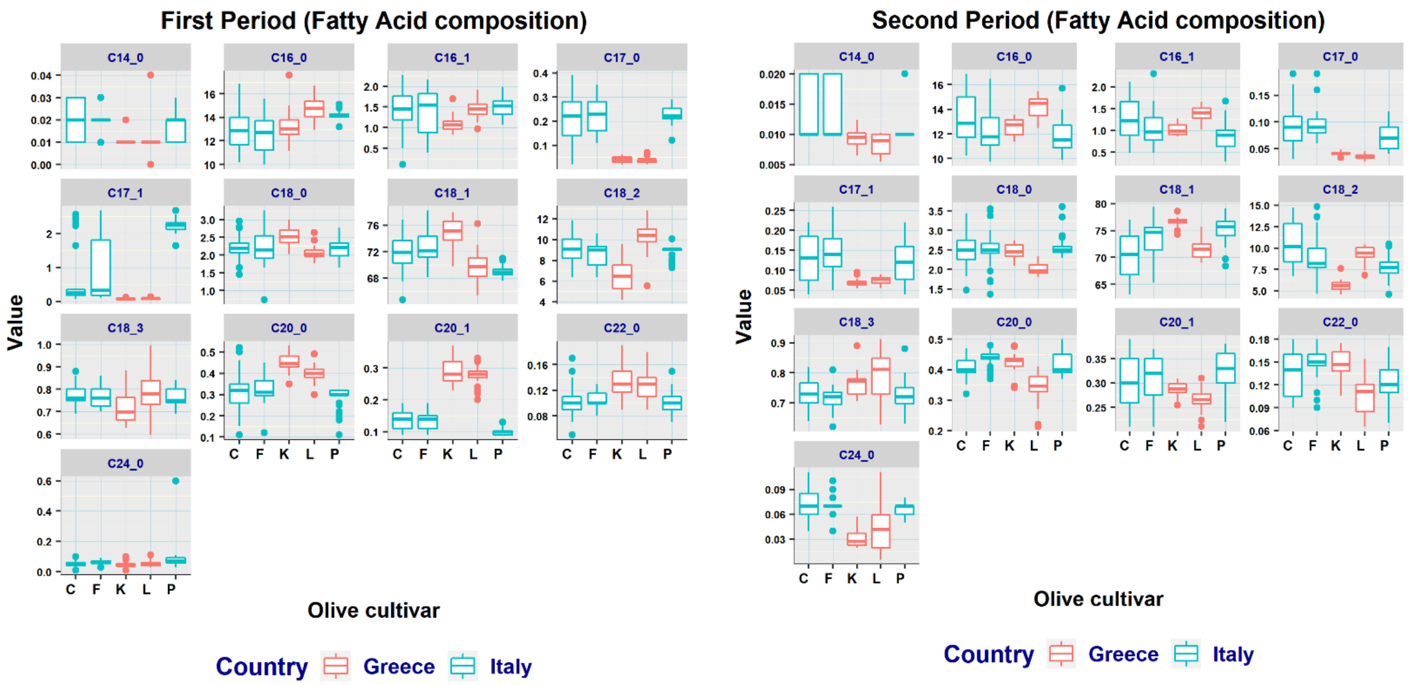

Figure 2. Box plot description of the fatty acid variables of the first and second sampling periods. C, F, K, L, P stand for Coratina, Favolosa, Koroneiki, Lianolia and Peranzana, respectively. Fatty acids are expressed in %.

3.2 Principal Components

We applied PCA to identify potential trends between the examined 28 olive oil compounds and olive varieties. We found eight significant components with values above the mean percentage of explained variance of all 28 components (Figure 3) and ten significant components when considering year separately (not shown). The first two components in the first and second years explained about 44% and 33% of the total variance, respectively, revealing trends in variables. When samples from the two years were pooled, about 37% of the total variance was explained. In the first year the Greek olive varieties were totally separable

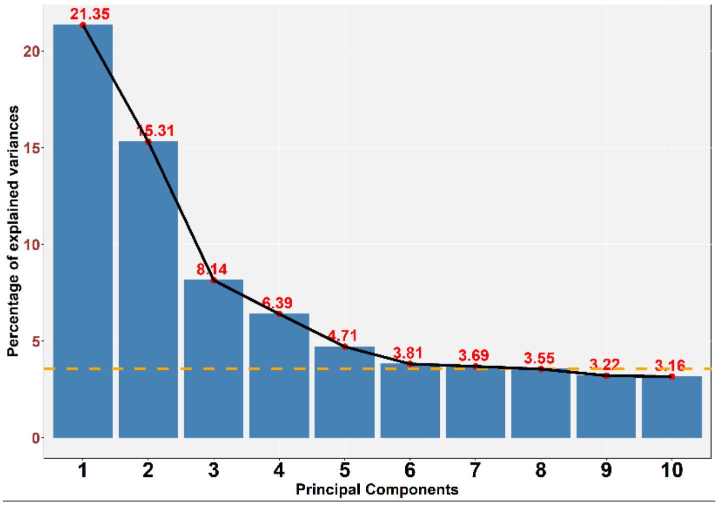

Figure 3 Scree plot showing the ten most important components in decreasing order of percentage of explained variance (red labels). Orange dashed line show the average percentage of explained variance of the 28 principal components.

from each other or from the Italian varieties (Figure 4). In contrast the Italian varieties showed extensive overlapping to each other. In the second year the Greek varieties were still totally separable from each other and from the Italian varieties, although a partial overlapping was observed in a small fraction of the Greek "Koroneiki" variety and the Italian varieties, which still were inseparable. The picture was more complicated when the two-year samples were pooled.

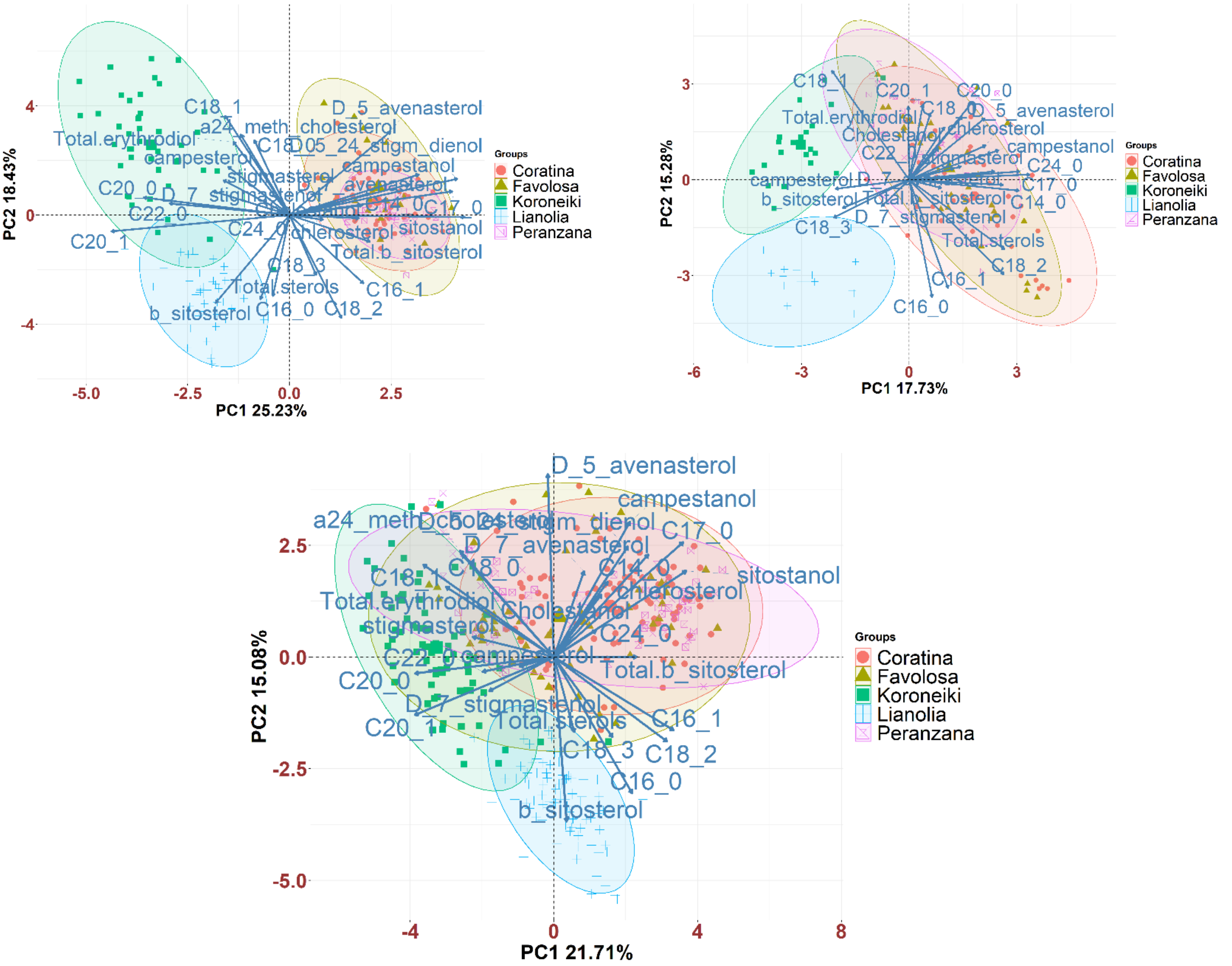

Figure 4. PCA analysis of the first two principal components for the first (upper left), second (upper right) and both periods (center).In the first period about 44% of the total variance is explained. Italian varieties are totally separable from Greek varieties, Greek varieties are almost totally separable from each other, and Italian varieties overlap. In the second period about 33% of the variance is explained and the same patterns are observed with the difference that Italian varieties overlap slightly with the Greek variety Koroneiki. When samples from both periods are pooled 37% of the total variance is explained. The discrimination between the Greek varieties is conserved, but an overlapping between the Italian varieties and the Koroneiki is observed.

In that case, although the Greek varieties were almost totally separable showing a small overlap, the three still inseparable Italian varieties showed a substantial overlay with the Greek variety "Koroneiki". These results suggest that the Greek varieties have innate differences in their olive oil composition, which are preserved from year to year (at least two years of the study period). It is also expected that most olive oil chemical compounds could vary within and between sampling periods and this variation could be easily detected by classical statistical approaches (e.g. ANOVA, Regression, non-parametric tests, etc.). In contrast, although differences of the chemical composition between the Italian and Greek varieties could be detected by classical statistics between and within sampling years, such differences could require a high number of samples to discriminate the overlapping variances. It is understandable that these suggestions concern the observed explained variance of 33-45% and more work is needed to explore the patterns of the rest principal component combinations.

3.3. Regression, multiple paired comparisons and non-parametric tests

We employed non-linear regressions to examine potential effect of the “Variety” or the “Sampling period” factors on each of the 28 olive oil compounds. We further examined paired significances of all combination of each factor levels, by using the “emmeans” library in “R”. These results are presented in the accompanying excel file named “***Olivnet_Regression and Comparisons.xlsx***”. As some of the required assumptions and particularly normality and equality of variances were non-met in several tests, all factors were tested by non-parametric omnibus tests. The Wilcoxin test (not shown) was used for factors with two levels and the Kruskal-Wallis test was used for factors with more than two levels and these results are presented in the accompanying file named “***Olivenet_Three_years_Kruskal_tests.xlsx***”. As these tests did not provide information on the significant pairs, a careful comparison between the results of the non-parametric and the corresponding parametric (regression) test is recommended for examining significances of a factors’ s levels on a chemical compound of particular interest.

Considering variation of olive oil compounds between years our analysis showed that the two Greek varieties (e.g. “cv. Koroneiki”, “cv. Lianolia”) were the most stable in a three year period, having 18 and 15 out of 28 chemical compounds, without significant differences (p>0.05), respectively. In contrast, the Italian varieties "Peranzana","Coratina" and "Favolosa" had six, seven and eight olive oil compounds with non-significant differences between years, respectively. The non-significant olive oil compounds for each variety are shown in Table 4. Eleven chemical compounds were common in the Greek varieties, six were common between Favolosa and Coratina, two were common between “Coratina” and “Peranzana”, and two were common between “Favolosa” and “Peranzana”. Four chemical compounds were common in four varieties, but none was common in all five varieties (Table 4). This results show that Koroneiki followed by Lianolia might express innate characteristics consistency among years most likely reflecting a constancy in their profile.

Considering olive oil compounds variation between varieties in each country per sampling period the results are presented in the accompanying file “**Between_Varieties_in_each_Country_Kruskal_Wallis_tests.xlsx”.** We found that the “Koroneiki” and “Lianolia” differed the most, having 24 out 28 olive oil compounds with significant differences in either year. In contrast, the Italian varieties showed adequate consistency in olive oil compounds variation during the two periods. However, this stability was not constant between the harvesting periods as the three varieties expressed extensive variation. Thus, in the first year the total number of olive oil compounds with at least one significance for the three Italian varieties was eight, while the same number in the second year was 20. These results suggest an innate difference in chemical characteristics expression between the two Greek varieties and a potential inhered similarity between at least some of the Italian varieties, affected however, by environmental or other related factors.

Table 4 Non-significant differences in the examined chemical compounds (sterols & fatty acid) of olive samples collected in a three year (Koroneiki, Lianolia) or a two-year period (Coratina, Peranzana, Favoloza). Blue squares correspond to no significant differences of the corresponding chemical compounds (p>0.05 among years)

Sterols	**Koroneiki**	**Lianolia**	**Favolosa**	**Coratina**	**Peranzana**
24_meth_cholesterol			■		
b_sitosterol	■	■			■
campestanol	■		■	■	■
campesterol	■	■			■
chlerosterol					■
Cholestanol		■	■	■	■
D_5_24_stigm_dienol	■	■			
D_5_avenasterol	■	■	■	■	
D_7_avenasterol	■	■	■	■	
D_7_stigmastenol	■				
sitostanol	■	■			
stigmasterol	■	■			
Total.b_sitosterol			■		
Total.erythrodiol	■				
Total.sterols	■	■			
C14_0	■				
C16_0			■	■	
C16_1	■	■		■	
C17_0	■	■			
C17_1	■	■			
C18_0	■				
C18_1			■	■	
C18_2			■		
C18_3		■			
C20_0					
C20_1	■				
C22_0	■				
C24_0		■			■

4. Machine learning

4.1. Dataset preparation

Supervised machine learning is an algorithmic structure (model) applied in a computer (machine) simulating human intelligence, which allows the gradual realization, through iterated self-regulating / self-adjusting processes, of the real parameters of an unknown function, which is then used to classify items, on characteristics on which the model has already been trained (supervised). Several approaches could be used for model training depending on the type of data that could be continuous or discrete variables.

One of the most difficult issues and challenging tasks is to identify patterns for olive oils geographical and botanical classification (Gonzalez-Fernandez et.al 20218, Jimenez-Carvelo, 2019).. In this work we used the olive dataset, obtained through the period of project implementation to train a model on the available records of olive oil chemical compounds paired to the corresponding olive varieties. We used decision trees as the predictive model and specifically extreme-boosted trees implemented by programming the XGBoost library in “R”.

Development of the model follows an hierarchical process, starting with the training and testing phase, which is applied in the randomly split dataset, progressing to the evaluation of the model by use of the so called confusing matrix, and the, necessary, if required, corrective approaches, ending up with the evaluation parameters of the determined model which could then be used in real world applications.

4.2. Training and Testing

We randomly split the dataset in training and testing subsets each holding about 80% and 20% of the records, respectively. The two subsets were processed and transformed in the appropriate matrix forms as required by the program. The training set was then fed to the training algorithm using the “softprob” as a loss function, and several sets of parameters were tested and optimized until the desirable minimization of the loss function was obtained, paying attention to avoid overfitting (e.g. high performance in the training subset but not in the testing subset or in real world data).

4.3. Confusion Matrix

The confusion matrix of existing (known) to predicted varieties as resulted by the trained model is shown in Table 5. The Italian varieties were perfectly separated from the Greek varieties with the only exception of “Koroneiki”, which in one case was mistaken to “Coratina”. The two Greek varieties, “Koroneiki” and “Lianolia”, were almost perfectly classified, with only one mistake where a “Koroneiki” was taken for “Lianolia”. The Italian varieties were correctly classified in high rates with some mistakes though. Thus “Coratina” was mistaken for “Favolosa” or “Peranzana” in five and two cases, respectively; “Favolosa” was mistaken for “Coratina” and

Table 5. Confusion matrix of known (Reference) vs predicted (Prediction) varieties resulted from model application on the testing subset.

		Reference				
		Coratina	Favolosa	Koroneiki	Lianolia	Peranzana
Prediction	Coratina	29	5	0	0	2
	Favolosa	3	13	0	0	1
	Koroneiki	1	0	24	0	0
	Lianolia	0	0	1	33	0
	Peranzana	3	3	0	0	20

"Peranzana" in three and one case, and "Peranzana" was mistaken for "Coratina" and "Favolosa" in three cases each. Overall, these results show a high performance of the trained model to classify varieties based on olive oil chemical characteristics (sterolic and fatty acid profile), even in cases were overlapping of the training variables is extensive (i.e. Italian varieties). It is expected that training with a larger dataset with more samples per harvesting year for each variety, the classification ability of the model will substantially improve.

4.4. Performance parameters

The obtained performance parameters, corresponding to the confusing matrix obtained by the trained model are show in Table 6. The most important features are Sensitivity, Specificity and Accuracy. Sensitivity is the ability of a model to correctly predict the true positives as positives, Specificity is the ability of predicting the true negatives as negatives and Accuracy is the rate of correct decisions. As our case was a multiclass classification problem, involving five prediction categories (as opposed to binary prediction of dichotomous variables), each category had its own performance parameters. The Greek varieties had high Sensitivity and Selectivity rates (most near 0.99) (Table 6) indicating almost perfect prediction among the five varieties. These rates corresponded to high accuracy rates of 0 .98 and 0.99 for the Koroneiki and Lianolia varieties indicating that in 100 samples 98 and 99 will be correctly classified.

Table 6 Performance parameters obtained from the confusion matrix.

Testing Dataset	VARIETY				
	Coratina	Favolosa	Koroneiki	Lianolia	Peranzana
Sensitivity	0.8056	0.619	0.96	1	0.8696
Specificity	0.9314	0.9658	0.9912	0.9905	0.9478
Pos Pred Value	0.8056	0.7647	0.96	0.9706	0.7692
Neg Pred Value	0.9314	0.9339	0.9912	1	0.9732
Balanced Accuracy	0.8685	0.7924	0.9756	0.9952	0.9087

The performance parameters of the Italian varieties were also excellent showing high rates mainly in Specificity. One exception, concerning Specificity was Favolosa with a low true positives rate of 0.62, most possibly resulting from the very small sample used for training. As the three Italian varieties had highly overlapping values a larger sample taken during each sampling period is expected to hugely improve performance of all varieties.

5. CONCLUSIONS

Olive oil samples of five monovarietal cultivars collected from several cloze locations in the northwestern part of Greece (cv. Koroneiki, cv. Lianolia) and the southern peninsular section of Italy (cv. Coratina, cv. Peranzane, cv. Favoloza) were analyzed for their chemical composition (sterolic and fatty acid profile) for three consecutive years in order to investigate the potential of developing a novel classification model capable of olive cultivar prediction.

Box plot analysis showed similarity patterns in some Italian varieties and analogous patterns in the two Greek varieties in terms of expression levels of the examined olive oil compounds which preserved across the years. Further exploration by PCA indicated extensive overlap among the three Italian varieties, and an innate difference between the Greek varieties. More detailed work in which varieties were compared by parametric and non-parametric tests, showed that the two Greek varieties showed considerably stability (non-significant differences) in most of the expressed olive oil chemical compounds across three cultivation periods whereas the Italian varieties showed variation (significant differences) in many of the examined olive oil compounds during two cultivation periods.

Similar statistical analysis showed considerable differences of most olive oil compounds between the two Greek varieties and these differences were preserved across the cultivation periods, suggesting innate distances (dissimilarities) between “Koroneiki” and “Lianolia”. In contrast more similarities in olive oil compounds were observed among the three Italian varieties, supporting a close genetic relationship among these varieties. In any case these similarities varied between the cultivation years suggesting the influence of other factors such as environment, agronomic involved in olive fruit physiology.

The final crucial step concerned the possibility of classifying olive oil samples of different varieties by training an artificial intelligence model with the utilization of the same olive dataset. The developed XGBoost model showed high ability in botanical discrimination, with the Greek varieties showing the highest performance followed by most Italian varieties with highly successful scores. Thus, these results are highly suggestive for incorporating machine learning technologies in olive oil varietal authentication.

Further in-depth research enriched with a higher number of samples and for more cultivation years will enable the improvement of the model. As this work belongs to the first systematic attempt focusing on botanical classification (Tahir, H.E.,2022) with the use of machine learning model (xgboost algorithm), further investigation is under way which will possible increase flexibility by selecting the least required number of the most informative olive oil compounds on the examined dataset

Finally this study can also contribute in the future to the establishment of a continuously enriched “Authentic Olive Network” of indigenous, local, and unexploited monovarietal olive oils while another promising extension of this work, with expected applications in analogous fields of the food industry (e.g. cheese) could be the training of artificial intelligence models for the detection of olive oil adulteration.

6. REFERENCES

Alboukadel Kassambara; Fabian Mundt (2020): factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Available online at https://CRAN.R-project.org/package=factoextra.

Daniel Ludecke; Mattan S. Ben-Shachar; Indrajeet Patil; Philip Waggoner; Dominique Makowski (2021): performance: An R Package for Assessment, Comparison and Testing of Statistical Models. In JOSS 6 (60), p. 3139. DOI: 10.21105/joss.03139.

Commission Regulation (EEC). No. 2568/91 of 14 July 1991 on the characteristics of olive oil and olive residue oil and on the relevant methods of analysis. Off. J. Eur. Union 1991, L208, 1–8. https://eur-lex.europa.eu/homepage.html

Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., Lacoste, F., Brereton, P., Garcia-Gonzalez, D. L., Moreda, W., & Gallina Toschi, T. (2020). Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future. Trends in Food Science &Technology, 105, 483–493. https://doi.org/ 10.1016/j.tifs.2019.02.025

Gonzalez-Fernandez, I., Iglesias-Otero, M. A. , Esteki, M., Moldes, O. A., Mejuto, J. C. & Simal-Gandara, J. (2019). A critical review on the use of artificial neural networks in olive oil production, characterization and authentication, Critical Reviews in Food Science and Nutrition, 59(12), 1913-1926, https://doi.org/10.1080/10408398.2018.1433628

Hadley Wickham (2016): ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York. Available online at https://ggplot2.tidyverse.org.

Hadley Wickham (2021): tidyr: Tidy Messy Data. Available online at https://CRAN.R-project.org/package=tidyr.

Hadley Wickham; Romain Francois; Lionel Henry; Kirill Muller (2021): dplyr: A Grammar of Data Manipulation. Available online at https://CRAN.R-project.org/package=dplyr.

Jimenez-Carvelo, A. M., Gonzalez-Casado, A., Bagur-Gonzalez, M. G., & Cuadros- Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review. Food Research International, 122, 25–39. https://doi.org/10.1016/j.foodres.2019.03.063

Lozano-Castellón, J., López-Yerena, A., Domínguez-López, I., Siscart-Serra, A., Fraga, N., Sámano, S., López-Sabater, C., Lamuela-Raventós, R. M., Vallverdú-Queralt, A., & Pérez, M. (2022). Extra virgin olive oil: A comprehensive review of efforts to ensure its authenticity, traceability, and safety. Comprehensive reviews in food science and food safety, 21(3), 2639–2664. https://doi.org/10.1111/1541-4337.12949

Max Kuhn (2020): caret: Classification and Regression Training. Available online at https://CRAN.R-project.org/package=caret.

Philipp Schauberger; Alexander Walker (2021): openxlsx: Read, Write and Edit xlsx Files.

R Core Team (2018): R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available online at https://www.R-project.org/.

Russell V. Lenth (2021): emmeans: Estimated Marginal Means, aka Least-Squares Means. Available online at https://CRAN.R-project.org/package=emmeans.

Sébastien Lê; Julie Josse; François Husson (2008): FactoMineR: A Package for Multivariate Analysis. In J. Stat. Soft. 25 (1), pp. 1–18. DOI: 10.18637/jss.v025.i01.

Stef van Buuren; Karin Groothuis-Oudshoorn (2011): mice: Multivariate Imputation by Chained Equations in R. In J. Stat. Soft. 45 (3), pp. 1–67. DOI: 10.18637/jss.v045.i03.

Tianqi Chen; Tong He; Michael Benesty; Vadim Khotilovich; Yuan Tang; Hyunsu Cho et al. (2021): xgboost: Extreme Gradient Boosting. Available online at https://CRAN.R-project.org/package=xgboost.

Tahir, H.E., Arslan, M., Mahunu, G.K., Mariod, A.A., Hashim, S. B. H., Xiaobo, Z., Jiyong, S., El-Seedi, H., Musa. T.H. (2022). The use of analytical techniques coupled with chemometrics for tracing the geographical origin of oils: A systematic review (2013–2020), Food Chemistry, 366, https://doi.org/10.1016/j.foodchem.2021.130633

7. SUPPLEMENTARY MATERIAL

1. ***Olivnet_Regression and Comparisons.xlsx***
2. ***Olivenet_Three_years_Kruskal_tests.xlsx***
3. **Between_Varieties_in_each_Country_Kruskal_Wallis_tests.xlsx**