Incorporating Random Tree Statistical Learning Classifier to Authenticate PDO Kalamata Olive Oil Blended with Aigialeia Olive Oil from the Geographic Region of Peloponnese in Greece

THEODOROS ANAGNOSTOPOULOS[1,2,3], IOAKIM SPILIOPOULOS[2], CHARA KOTTARA[3]
[1]Department of Business Administration,
University of the Aegean,
Michalon 8, Chios 821 00,
GREECE

[2]Department of Food Science and Technology,
University of Peloponnese,
Antikalamos, 24100 Kalamata,
GREECE

[3]Department of Business Administration,
University of West Attica
250 Thivon & P. Ralli Str, Egaleo, 12241 Athens,
GREECE

Abstract: Authentication of Protected Designation of Origin (PDO) Kalamata olive oil is required to assess its quality in the marketplace compared with other olive oil varieties. Concretely, Kalamata is located in southern Greece in the geographic county of Messenia, which is part of the geographic region of Peloponnese and is famous for its extra virgin olive oil produced from the Koroneiki olive variety. Intuitively, PDO Kalamata olive oil, established by Council regulation (EC) No 510/2006, owes its quality and special characteristics to the geographical environment, olive tree variety, and human factor. However, authentication of the PDO Kalamata olive oil is a challenging task when it is blended with other olive oil varieties, such as the Aigialeia olive oil variety that is cultivated in the geographic county of Achaia, which is also located in the geographic region of Peloponnese. Subsequently, the PDO Kalamata olive oil authentication process is achieved by adopting the potentiality of certain statistical machine learning models. Specifically, in this paper, a random tree classification model to authenticate PDO Kalamata olive oil when it is blended with olive oil from Aigialeia. Concretely, the adopted classification model authenticates the quality of the PDO Kalamata olive oil variety based on synchronous excitation-emission fluorescence (SyEE) spectroscopy applied to certain olive oil data samples. Experiments performed evaluate the efficiency of the adopted random tree statistical learning classifier. Intuitively, the observed results promise to define the originality and authentication of the PDO Kalamata extra virgin olive oil by exploiting its unique quality characteristics.

Key-Words: - Olive oil authentication, blended olive oil, synchronous emission-excitation, fluorescence spectroscopy, statistical learning, binary classification, random tree model evaluation.

Received: June 15, 2024. Revised: December 14, 2024. Accepted: February 9, 2025. Published: March 14, 2025.

1 Introduction

Protected Designation of Origin (PDO) Kalamata olive oil is considered a high-quality olive oil, which is cultivated in the province of Messenia in the geographic region of Peloponnese in southern Greece, [1]. PDO Kalamata olive oil, established by Council regulation (EC) No 510/2006, owes its quality and special characteristics to the geographical environment, olive tree variety, and human factor. There is a need to authenticate the originality of such an olive oil before providing it to local and global marketplace, [2]. However, there is the risk of blending PDO Kalamata olive oil with other olive oil varieties of olive oils from other areas, such as the case of the Aigialeia olive oil that is cultivated in the geographic county of Achaia in the geographic region Peloponnese. To perform authentication of the PDO Kalamata olive oil requires the adoption of certain statistical learning classification models, [3]. Such machine learning models are able to assess the potentiality of specific data produced by performing detailed experimental measurements of synchronous excitation-emission fluorescence spectroscopy samples, [4].

Concretely, to protect high-quality olive oil and prevent adulteration with low-quality oils, international governmental agencies like the: (1) Codex Alimentarius, (2) International Olive Council, and (3) European Commission, have developed certain standards to regulate olive oil by establishing a set of physical, chemical, and organoleptic characteristics, [5]. Intuitively, traditional chemical methods to ensure high-quality olive oil are focused on the authentication and quantification of pre-defined compounds or data classes of compounds of olive oil samples according to the regulations of the above-mentioned international governmental agencies. However, such methods are time-consuming and demand expensive apparatus.

Contemporary research focuses on non-targeted analysis, which has attracted much research attention. Such an approach incorporates screening of olive oil without any prior knowledge of chemical composition. Concretely, in this approach, certain data analytics methods are used that produce a signal that is affected by all the compounds (i.e., metabolites) present in provided olive oil samples. This approach shortens the analytical process. However, there is a need for a vast number of data sources, which are required to perform data analytics based on statistical machine learning models, [6].

In this paper a random tree statistical learning classifier is adopted to assess the originality and authentication of PDO Kalamata olive oil variety, [7]. Such a classification model is able to distinguish PDO Kalamata high-quality olive oil when it is blended with olive oil from Aigialeia in certain quantities. Aigialeia is a municipality and geographical area in northern Peloponnese with significant olive oil production. Concretely, blended olive oil is produced by blending 60 percent of PDO Kalamata high-quality olive oil with 40 percent of Aigialeia low-quality olive oil. Specifically, the proposed classification model assesses the potentiality of data sources produced by measuring synchronous excitation-emission fluorescence spectroscopy olive oil samples. Certain evaluation methods and metrics are incorporated during certain experiments to observe robust results. Concretely, results are promising in defining the originality and authentication of PDO Kalamata olive oil.

The rest of the paper is organized as follows. Section 2 presents prior work performed in the research area. Section 3 defines the proposed data model. In Section 4, evaluation method and metrics are adopted. In Section 5, experiments are performed, and results are presented. Section 6 discusses the strengths and the weaknesses of the proposed research study, while Section 7 concludes the paper and indicates possible future work.

2 Prior Work

Assessing live oil originality and authentication is achieved by incorporating synchronous excitation-emission fluorescence spectroscopy. Concretely, coupled near-infrared (NIR) and fluorescence excitation-emission matrix spectroscopies are adopted as chemometric tools to evaluate the authentication of certain virgin olives, [8]. Intuitively, Fourier-transform infrared (FTIR), visible-near-infrared (Vis-NIR) and excitation-emission matrix fluorescence spectroscopy (EEMs) have been used to detect olive oil adulteration. The data are analyzed by principal component analysis (PCA)/ multi-way-PCA analysis, [9]. Authentication of extra virgin olive oil is achieved by incorporating a fusion of fluorescence with near-infrared spectroscopy data sources assessing the potentiality of a multi-block Data Driven-Soft Independent Modeling of Class Analogies (DD-SIMCA) method, [10]. Originality and authentication of extra virgin olive oils were assessed by a detailed pigment profile coupled with spectroscopic techniques, [11]. Multivariate data analytics are also applied on absorption and fluorescence spectroscopic data to define a method to determine extra virgin olive oil adulteration, [12].

A rapid detection method is incorporated to define the authenticity and adulteration of olive oil through the excitation-emission matrix of fluorescence, [13]. A microwave dielectric spectroscopy technique is used to identify the quality of extra virgin olive oil, which assesses the aid provided by certain feedforward neural networks, [14]. Fluorescence spectroscopy, which is based on a machine learning model, is incorporated to assess the quality of extra virgin olive oil during the aging process, [15]. Identification of originality and authentication of extra virgin olive oil is achieved by using miniaturized near infrared instruments according to a rapid non-destructive process, [16]. Feasible Identification, authentication as well as adulteration of olive oil was based on FT-NIR spectroscopy incorporating certain machine learning models, which assesses the potentiality of the feature selection process applied to the provided data sources, [17].

Nuclear Magnetic Resonance spectroscopy supported by vector machines was applied for rapid screening of olive oil originality and authentication, [18]. A support vector machine classification algorithm that exploits a certain pigment composition technique is incorporated to identify and authenticate extra virgin olive oil, which is adulterated with other edible oils, [19]. A machine learning model is used to evaluate olive oil adulteration, which is based on Fourier-transform mid-infrared (FT-MIR) as and near-infrared (FT-NIR) spectroscopies that are associated with specific multivariate analyses, [20]. Emphasis on non-invasive methods for assuring the originality and identification of olive oils is incorporated, which adopts certain machine learning inference models to provide authentication information, [21]. Laser Induced Breakdown Spectroscopy is used, assisted by machine learning algorithms, as an ultrafast methodology for olive oil classification in terms of the degree of acidity and geographical origin, [22]. A research approach that is based on the extra virgin olive oil cultivar and ripening stage is proposed, which examines the potential of applying machine learning algorithms to experimental data produced by low-frequency dielectric spectroscopy, [23].

A convolutional neural network is incorporated to perform quantitative statistical analysis on Raman spectroscopy data of blended olive oil, [24]. Classification of oil type and the quick detection of adulterated edible oils is achieved by specific statistical machine learning models, which exploit the potentiality of Raman spectroscopy, [25]. Vibrational spectroscopic analytical machine learning techniques are incorporated to define observed authentication trends of edible olive oils, [26]. Contemporary statistical machine learning models are also used to efficiently detect olive oil adulteration, [27]. A sensor system for food authentication was based on computer vision and pattern recognition using smartphone cameras to record the color of olive oil, [28]. A detailed analysis, which incorporates statistical deep machine learning models, is able to detect olive oil on the tree branches, thus exploiting possible agricultural practices, [29]. Stochastic analysis, which uses analytical methods of fluid-chromatography-mass spectrometry investigation, is applied to olive stones, fruits, and leaves to ensure the overall olive oil quality, [30].

Contemporary research in the area of olive oil originality has had a significant impact on several olive oil varieties. Concretely, current research effort focuses on identifying the authentication of PDO Kalamata high-quality olive oil variety when it is blended with Aigialeia olive oil producedt in the Aigialeia area thus preventing possible fraud in the local and global olive oil marketplace. To perform such olive oil originality identification, it is exploited by the potentiality of a random tree statistical machine learning classifier.

3 Data Model

Proposed research effort focuses on the definition of a data model able to input certain machine learning algorithms to provide efficient results. Pure PDO Kalamata and Aigialeia olive oil physical samples used to perform statistical data analytics were characterized by synchronous emission-excitation fluorescence (SyEE) spectra. Physical spectra are recorded on a Perkin Elmer LS55 spectrofluorometer using solution 1% w/v olive oil in n-hexane, where $\Delta\lambda$ (i.e., the difference between excitation and emission wavelength) is adjusted to 30 nm, [31]. The excitation and emission slit are tuned to 4 nm. The scan rate is 50nm/min. Each olive oil physical sample is measured triplicate using the new freshly prepared solution. Each measurement of an olive oil physical sample was statistically processed as a different physical sample of the same olive oil origin. Concretely, certain SyEE data sources from PDO Kalamata olive oil and Aigialeia olive oil are synthetically blended based on data analytics to produce a synthetic blended olive oil for further experimentation. Such synthetic blending indicates the experimental nature of the observed data that input the adopted statistical learning classifier. Intuitively, blended olive oil is produced by mixing 60 percent of PDO Kalamata high-quality olive oil with 40 percent of Aigialeia low-quality olive oil. Measurements were produced on physical samples of both classes (i.e., PDO Kalamata olive oil and blended olive oil), while focus is given on the potentiality to identify originality and authentication of PDO Kalamata olive oil.

3.1 Data Structure

Synchronous emission-excitation fluorescence spectra are examined based on certain predictive attributes values measured, which define one of the two possible class values. Concretely, the PDO Kalamata authentic olive oil class value is denoted as Class 0, while the blended olive oil class value is denoted as Class 1. Since provided data sources are divided into two separate classes (i.e., Class 0 and

Class 1), the studied predictive classification problem is considered as a binary classification problem. Total number of data samples incorporated in the current research effort are 58, where 29 are PDO Kalamata authentic olive oil samples. The remaining 29 data samples are forming the blended olive oil produced by mixing PDO Kalamata high-quality olive oil with Aigialeia olive oil. Provided data sources are limited, thus to treat this issue, there is a need to adopt certain evaluation methods and metrics.

3.1.1 PDO Kalamata Olive Oil

Data structure of authentic PDO Kalamata high-quality olive oil has certain predictive attributes' values. Intuitively, tocopherols are within a range of [52.411, 209.902] values, and phenolic compounds are defined within a range of [6.568, 13.785] values. Oxidation products of triglycerides are within a range of [0.174, 12.283] values, while oxidation products of tocopherols are defined within a range of [0.625, 8.618] values. Concretely, chlorophylls are within a range of [30.633, 62.382] values. PDO Kalamata authentic olive oil observed data structure is presented in Table 1.

Table 1. PDO Kalamata Olive Oil Data Structure

Predictive Attribute	Min Value	Max Value
Tocopherols	52.411	209.902
Phenolic compounds	6.568	13.785
Oxidation products of triglycerides	0.174	12.283
Oxidation products of tocopherols	0.625	8.618
Chlorophylls	30.633	62.382

3.1.2 Blended Olive Oil

Data structure of blended olive oil has also specific predictive attributes' values. Subsequently, tocopherols are defined within a range of [145.431, 150.184] values, phenolic compounds are within a range of [10.782, 15.533] values. Oxidation products of triglycerides are defined within a range of [8.098, 13.099] values, while oxidation products of tocopherols are within a range of [1.686, 6.435] values. Concretely, chlorophylls are defined within a range of [64.531, 68.283] values. Blended olive oil provided data structure is presented in Table 2.

Table 2. Blended Olive Oil Data Structure

Predictive Attribute	Min Value	Max Value
Tocopherols	145.431	150.184
Phenolic compounds	10.782	15.533
Oxidation products of triglycerides	8.098	13.099
Oxidation products of tocopherols	1.686	6.435
Chlorophylls	64.531	68.283

3.2 Provided Data Visualization

Visualization of input data provides the feasibility to understand in deep detail the characteristics of the provided data sources' distribution and visual behavior. Since, in the current research study, there are 2 data classes, it holds that Class 0 is assigned to PDO Kalamata olive oil, and Class 1 is assigned to blended olive oil. Visualization of adopted data classes are provided in Figure 1. Specifically, in the lower left corner of the figure it can be observed with blue signs the distribution of Class 0, i.e., PDO Kalamata olive oil, while in the upper right corner of the figure it can be observed in red signs the distribution of the input Class 1, i.e., blended olive oil.

4 Evaluation Parameters

The performance of the proposed random tree statistical machine learning model is assessed by incorporating specific evaluation methods and metrics. Such evaluation parameters are used to perform certain experiments and observe the output results.

4.1 Evaluation method

Evaluation of the random tree machine learning model is performed by incorporating specific evaluation methods. In current research effort a fundamental evaluation method is used, due to its simplicity and optimum output results, which is the 10-fold cross-validation, [32].

Fig. 1: Data visualization, i.e., blue signs of initial Class 0, and red signs of initial Class 1

Concretely, the adopted evaluation method divides the input dataset into 10 equal sized parts and then in a certain loop incorporates the first 9 parts to train the statistical learning classification algorithm and the remaining 1 to test the classifier. Described analytical process is repeated until all the parts are used for training and testing. The adopted evaluation method is used in the proposed machine learning analytics since it provides efficient output results based on certain input data, which are able to explain the observed data source's predictive analytics behavior.

4.2 Evaluation Metrics

Provided the adopted 10-fold cross-validation method that is used to support the experimental setup, there is a need to incorporate specific evaluation metrics. Such evaluation metrics are (1) prediction accuracy, (2) correctly classified instances, and (3) confusion matrix that can exploit the potentiality of the proposed random tree statistical classification model.

4.2.1 Prediction Accuracy

The efficiency of the proposed random tree statistical machine learning algorithm is assessed by using the prediction accuracy evaluation metric, $a \in [0, 1]$, which is defined as presented in the following mathematical equation, (1):

$$a = \frac{tr_{pos}+tr_{neg}}{tr_{pos}+fl_{pos}+tr_{neg}+fl_{neg}} \qquad (1)$$

where, tr_{pos}, is the number of instances, which are classified correct as positives, and, tr_{neg}, is the number of instances, which are classified correct as negatives. In addition, fl_{pos}, is the number of the instances, which are classified false as positives, and, fl_{neg}, is the number of the instances that are classified false as negatives. A low value of prediction accuracy a means a weak classification algorithm while a high value of a indicates an effective statistical machine learning model. Intuitively, experimental assessment based on the defined statistical quantities of: (1) tr_{pos}, (2) tr_{neg}, (3) fl_{pos}, and (4) fl_{neg}, which compose the prediction accuracy evaluation metric's experimental value, achieve to express the data sources' dynamics and explain the observed optimal output results.

4.2.2 Correctly Classified Instances

To assess a machine learning algorithm effectiveness, it is common to express prediction accuracy numerical value as a percentage thus observed output results being more easily interpreted and presented. Subsequently, it is used the term of correctly classified instances, $c \in [0\%, 100\%]$, which is defined according to the following mathematical, analytical equation, (2):

$$c = \alpha\% \qquad (2)$$

where, a value, which is close to 0% means that the classification algorithm is not efficient, while a value that is close to 100% indicates that the statistical machine learning model is able to classify instances optimally.

4.2.3 Confusion Matrix

Adopted statistical classification machine learning random tree algorithm is also evaluated with the confusion matrix evaluation metric. Confusion matrix is a special form of matrix, which in the case of a binary classification of 2 classes, (i.e., Class 0: PDO Kalamata olive oil, and Class 1: blended olive oil) has the following analytical form, as described in Table 3.

Where, "A" quantity depicts the number of Class 0 instances, which are classified correctly as instances of Class 0. "B" quantity depicts the number of Class 0 instances, which are falsely classified as instances of Class 1. "C" quantity depicts the number of Class 1 instances, which are falsely classified as instances of Class 0, and the "D" quantity depicts the number of Class 1 instances, which are correctly classified as instances of Class 1. An adopted machine learning classification algorithm is considered effective if it maximizes the elements of the main diagonal of the confusion matrix (i.e., "A", and "D") and minimizes the other elements (i.e., "B", and "C"). A confusion matrix is incorporated in statistical learning evaluation methodology to indicate effectiveness and explain in deep detail the statistical nature of output results computed by the prediction accuracy evaluation metric.

Table 3. Confusion matrix

Class 0	Class 1	← Classified as
A	B	Class 0
C	**D**	Class 1

5 Experiments and Results

The adopted data model contains input data either from PDO Kalamata olive oil or from blended olive oil samples. Such data sources are used to perform certain experiments to assess the efficiency of the proposed random tree machine-learning classification model to distinguish the originality of the authentic PDO Kalamata high-quality olive oil from blended olive oil. A certain experimental setup is required to perform analytics during the experimental phase with specific evaluation methods and evaluation metrics to observe the output results.

5.1 Experimental Setup

Certain parameters are used to set up the experimental data analytics process. Intuitively, it is defined as the number of the observed classes which is assigned to each data sample instance. Subsequently, predictive attributes used to describe a certain class are defined based on input data sources accordingly. Concretely, a specific statistical machine learning algorithm is proposed to perform the experiments and observe the output results.

5.1.1 Binary Classification

Due to the fact that the number of classes is 2 this classification data analytics process is characterized as a binary classification problem. Concretely, the 2 classes are defined as follows: (1) Class 0 is assigned to PDO Kalamata olive oil, and (2) Class 1 is assigned to blended olive oil. Intuitively, the number of predictive attributes is settled to be 5, which are defined as follows: (1) 1st predictive attribute: 'tocopherols', (2) 2nd predictive attribute: 'phenolic compounds', (3) 3rd predictive attribute: 'oxidation products of triglycerides', (4) 4th predictive attribute: 'oxidation products of tocopherols', and (5) 5th predictive attribute: 'chlorophylls'. Subsequently, the total number of data sample instances is 58, which have the following distribution per class: (1) 29 samples from Class 0, i.e., PDO Kalamata olive oil, and (2) 29 samples from Class 1, i.e., blended olive oil.

5.1.2 Random Tree Machine Learning Model

Machine model selection is a process required to find the optimal classification algorithm that is efficient in the examined binary classification problem. Several experiments were performed with certain statistical machine learning classification algorithms available in the Weka machine learning software, [33]. Concretely, the adopted statistical learning algorithm, which has optimal predictive accuracy behavior, emerged to be the Random Tree Machine Learning Model, thus, it is incorporated for further experimentation to observe the output results of the current research effort.

5.2 Derived Results

Evaluation of the adopted machine learning classifier is achieved by using an experimental phase where a certain evaluation method is defined (i.e., 10-fold cross-validation). Intuitively, specific evaluation metrics are incorporated to assess the effectiveness of the proposed machine learning algorithm, which in this case is the random tree machine learning model. Subsequently, based on specific evaluation parameters, certain output results are observed, which define the efficiency of the adopted experimental setup incorporated in the current research paper. Specifically, to assess the output results and be able to explain the research paper's findings, it is significant to incorporate the defined evaluation method and metrics. Such knowledge would reveal the inherent complexity that exists in the provided input data with a special focus on observing optimal output results for the proposed machine learning classification model.

5.2.1 Observed Prediction Accuracy

The evaluation method used to assess the efficiency of the proposed machine learning binary classification random tree algorithm is 10-fold cross-validation. According to the adopted evaluation method, it is observed a prediction accuracy with a value of: $a = 0.9827$, which is a relatively high value for prediction accuracy that proves the proposed machine learning model is convenient for the examined binary classification problem. Subsequently, the relatively high value of the observed prediction accuracy indicates that the adopted machine learning model could be used for predicting similar new unseen olive oil data instances in future research that might extend the potentiality of the current research paper to another alternative geographic region in Greece.

5.2.2 Observed Correctly Classified Instances

Based on the adopted evaluation method of 10-fold cross-validation, it holds that the observed correctly classified instances have a value of: $c = 98.27\%$, which indicates that the proposed machine learning classification model constitutes an optimal selected choice for the examined binary classification problem.

5.2.3 Observed Binary Confusion Matrix

Binary confusion matrix results as derived from the experimental setup based on a 10-fold validation evaluation method for the examined binary classification problem. Derived results are presented in Table 4.

Table 4. Binary confusion matrix results

Class 0	Class 1	← Classified as
28	1	Class 0
0	**29**	Class 1

At this point of the current research effort, it can be observed that the majority of the classified instances are located in the main diagonal of Table

4. Concretely, the quantity of elements in the main diagonal depicts the significant number of certain instances, which are correctly classified during the experimentation phase. Intuitively, such an effective prediction behavior constitutes a strong classification model for the examined binary classification problem. Such an optimal binary confusion matrix enables the observation of output experimental results in great detail, thus being able to understand the efficiency of the proposed random tree statistical learning model for predicting PDO Kalamata olive oil to assure its originality and authenticity compared with the blended olive oil data sample instances.

5.3 Classified Data Visualization

Visualization of the output classified data is considered an effective tool to be able to assess visually the results of the classification process. Such feasibility provides a deep understanding of the resulting data characteristics and dynamic behavior. Visualization of classified data sources is provided in Figure 2. Concretely, in the lower left corner of the figure is presented in blue signs the distribution of Class 0, i.e., PDO Kalamata olive oil, which has correctly classified as Class 0. Continuously, in the upper right corner of the figure, it can be observed in red signs the correctly classified instances of Class 1, i.e., blended olive oil. However, a data sample instance of Class 0, i.e., red sign, is falsely classified as an instance of Class 1. This is presented as a sole blue sign in the red signs distribution in the upper right corner of the figure.

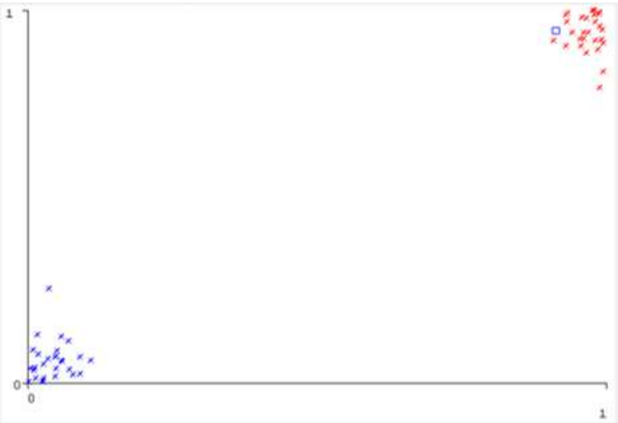

Fig. 2: Visualization of classification results, i.e., blue signs of classified Class 0 and red signs of classified Class 1

6 Conclusions and Future Work

The examined problem definition constitutes a binary classification problem of 2 discrete classes along with five separate predictive attributes per class. Concretely, there are a total of 58 sample instances, which are divided into 28 samples from PDO Kalamatas olive oil and another 28 samples produced by computational blending SyEE data of PDO Kalamata with the ones of olive oil from the geographical region of Aigialeia. The physical olive oil samples PDO Kalamata and Aigialeia are produced in the geographic region of Peloponnese in Greece. Specifically, 28 PDO Kalamata high-quality olive oil is further cultivated in the geographic county of Messenia in Peloponnese. Concretely, blended olive oil is produced by mixing PDO Kalamata high-quality olive oil with Aigialeia olive oil, which is cultivated in the geographic county of Achaia in the Peloponnese. It holds that the current research study has achieved high values of predicted output results based on certain evaluation methods and metrics, which indicate the robustness of the adopted evaluation parameters.

Future work should focus on updating current physical olive oil samples of both PDO Kalamata high-quality olive oil and blended olive oil with more new unseen samples, which can be incorporated to further validate the efficiency and robustness of the adopted random tree statistical machine learning classification algorithm. Such a methodological direction is able to form an effective predictive data analytics tool, which could be adopted by the research community to ensure originality and authentication of PDO Kalamata's high-quality olive oil Koroneiki variety. Concretely, research findings could be patented and further adopted by the local and/or global marketplace to provide a system of assuring the validation process of the current research effort, such as a possible collaboration with PDO Kalamata olive oil retail as well as wholesale companies, [34].

Acknowledgement:
The authors would like to acknowledge the Department of Food Science and Technology at the University of Peloponnese, Greece for providing real and synthetic blended olive oil data sources incorporated in the current research effort.

References:

[1] L. Trabelsi, B. Ncube, A. B. Hassena, M. Zouairi, F. B. Amar, and K. Gargouri, Comparative study of productive performance

of two olive oil cultivars Chemlali Sfax and Koroneiki under arid conditions, *South African Journal of Botany*, Vol. 154, 2023, pp. 356–364, https://doi.org/10.1016/j.sajb.2023.01.055.

[2] P. Rajak, A. Ganguly, S, Adhikary, and S. Bhattacharya, Internet of Things and smart sensors in agriculture: Scopes and challenges, *Journal of Agriculture and Food Research*, Vol. 14, 2023, pp. 1–13, https://doi.org/10.1016/j.jafr.2023.100776.

[3] M. E. Schiano, F. Sodano, C. Cassiano, E. Magli, S. Seccia, M. G, Rimoli, and S. Albrizio, Monitoring of seven pesticide residues by LC-MS/MS in extra virgin olive oil samples and risk assessment for consumers, *Food Chemistry*, Vol. 442, 2024, pp. 1–8, https://doi.org/10.1016/j.foodchem.2024.138498.

[4] J. Krause, H. Gruger, L. Gebauer, X. Zheng, J. Knobbe, T. Pgner, A. Kicherer, R. Gruna, T. Langle, and J. Beyerer, Smart Spectrometer–Embedded Optical Spectroscopy for Applications in Agriculture and Industry, *Sensors*, Vol. 21, 2021, pp. 1–18, https://doi.org/10.3390/s21134476.

[5] R. Aparicio, M. T. Morales, R. A. Ruiz, N. Tena, and D. L. G. González, Authenticity of olive oil: Mapping and comparing official methods and promising alternatives, *Food Research International*, Vol. 54, No. 2, 2013, pp. 2025–2038, https://doi.org/10.1016/j.foodres.2013.07.039.

[6] D. I. Ellis, H. Muhamadali, S. A. Haughey, C. T. Elliott, and R. Goodacre, Point–and–shoot: Rapid quantitative detection methods for on–site food fraud analysis–moving out of the laboratory and into the food supply chain, *Analytical Methods*, Vol. 7, No. 22, 2015, pp. 9375–9716, https://doi.org/10.1039/C5AY02048D.

[7] X. Miao, J. Ma, X. Miu, H. Zhang, Y. Geng, W. Hu, Y. Deng, and N. Li, Integrated transcriptome and proteome analysis the molecular mechanisms of nutritional quality in 'Chenggu-32' and 'Koroneiki' olives fruits (Olea europaea L.), *Journal of Plant Physiology*, Vol. 288, 2023, pp. 1–12, https://doi.org/10.1016/j.jplph.2023.154072.

[8] A. M. Jimenez-Carvelo, V. A. Lozano, and A. C. Olivieri, Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils, *Food Control*, Vol. 96, 2019, pp. 22–28, https://doi.org/10.1016/j.foodcont.2018.08.024.

[9] X. Meng, C. Yin, L. Yuan, Y. Zhang, Y. Ju, K. Xin, W. Chen, K. Lv, and L. Hu, Rapid detection of adulteration of olive oil with soybean oil combined with chemometrics by Fourier transform infrared, visible-near-infrared and excitation emission matrix fluorescence spectroscopy: A comparative study, *Food Chemistry*, Vol. 405, 2023, pp. 1–10. https://doi.org/10.1016/j.foodchem.2022.134828.

[10] V. A. Lozano, A. M. J. Carvelo, A. C. Olivieri, S. V. Kucheryavskiy, O. Y. Rodionova, and A. L. Pomerantsev, Authentication of Argentinean extra-virgin olive oils using three-way fluorescence and two-way near-infrared data fused with multi-block DD-SIMCA, *Food Chemistry*, Vol. 463, 2025, pp. 1–8, https://doi.org/10.1016/j.foodchem.2024.141127.

[11] O. Uncu, B. Ozen, and F. Tokatli, Authentication of Turkish olive oils by using detailed pigment profile and spectroscopic techniques, *Journal of the Science of Food and Agriculture*, Vol. 100, 2020, pp. 2153–2165, https://doi.org/10.1002/jsfa.10239.

[12] G. Stavrakakis, A. Philippidis, and M. Velegrakis, Application of Optical Spectroscopic Techniques and Multivariate Statistical Analysis as a Method of Determining the Percentage and Type of Adulteration of Extra Virgin Olive Oil, *Food Analytical Methods*, Vol. 15, 2022, pp. 285–293, https://doi.org/10.1007/s12161-021-02055-8.

[13] Y. Y. Yuan, S. T. Wang, J. Z. Wang, Q. Cheng, X. J. Wu, and D. M. Kong, Rapid detection of the authenticity and adulteration of sesame oil using excitation-emission matrix fluorescence and chemometric methods, *Food Control*, Vol. 112, 2020, pp. 1–9, https://doi.org/10.1016/j.foodcont.2020.107145.

[14] J. C. P. Alarcon, M. I. O. Souza, V. M. Pepino and B. H. V. Borges, Identification and Quantification of Common Adulterants in Extra Virgin Olive Oil Using Microwave Dielectric Spectroscopy Aided by Feedforward Neural Networks, *IEEE Sensors Journal*, Vol. 24, No. 19, 2024, pp. 29985–

29995, https://doi.org/10.1109/JSEN.2024.3448221.

[15] F. Venturini, S. Fluri, M. Mejari, M. Baumgartner, D. Piga, and U. Michelucci. Machine learning-enhanced fluorescence spectroscopy for the quality assessment of extra virgin olive oil during ageing, *Proceedings SPIE 12999, Optical Sensing and Detection VIII*, 129991F, 20 June 2024, Strasbourg, France, https://doi.org/10.1117/12.3016879.

[16] A. A. Cerezo, X. Yang, A. M. J. Carvelo, M. Pellegrino, A. F. Savino, and P. Berzaghi, Assessment of extra virgin olive oil quality by miniaturized near infrared instruments in a rapid and non-destructive procedure, *Food Chemistry*, Vol. 430, 2024, pp. 1–8, https://doi.org/10.1016/j.foodchem.2023.137043.

[17] L. S. Vieira, C. Assis, M. Queiroz, A. A. Neves, and A. F. Oliveira, Building robust models for identification of adulteration in olive oil using FT-NIR, PLS-DA and variable selection, *Food Chemistry*, Vol. 345, 2021, pp, 1–7, https://doi.org/10.1016/j.foodchem.2020.128866.

[18] X. Wang, G. Wang, X. Hou, and S. Nie, A Rapid Screening Approach for Authentication of Olive Oil and Classification of Binary Blends of Olive Oils Using Low-Field Nuclear Magnetic Resonance Spectra and Support Vector Machine, *Food Analytical Methods*, Vol. 13. No. 10, 2020, pp. 1894–1905, https://doi.org/10.1007/s12161-020-01799-z.

[19] C. H. Lu, B. Q. Li, Q. Jing, D. Pei, and X. Y. Huang, A classification and identification model of extra virgin olive oil adulterated with other edible oils based on pigment compositions and support vector machine. *Food Chemistry*, Vol. 420, 2023, pp. 1–9, https://doi.org/10.1016/j.foodchem.2023.136161.

[20] C. G. Pereira, A. I. N. Leite, J. Andrade, M. J. V. Bell, and V. Anjos, Evaluation of butter oil adulteration with soybean oil by FT-MIR and FT-NIR spectroscopies and multivariate analyses, *LWT*, Vol. 107, 2019, pp. 1–8, https://doi.org/10.1016/j.lwt.2019.02.072.

[21] E. E Okere, E. Arendse, H. Nieuwoudt, O. A. Fawole, W. J. Perold, and U. L. Opara, Non-invasive methods for predicting the quality of processed horticultural food products, with emphasis on dried powders, juices and oils: A review, *Foods*, Vol. 10, No. 12, 2021, pp. 1–31, https://doi.org/10.3390/foods10123061.

[22] O. Gazeli, E. Bellou, , FoodD. Stefas, S. Couris, Laser-based classification of olive oils assisted by machine learning, *Food Chemistry*, Vol. 302, 2020, pp. 1–7, https://doi.org/10.1016/j.foodchem.2019.125329.

[23] M. Rashvand, G. Altieri, A. Matera, F. Genovese, and G. C. D. Renzo, Potential of low frequency dielectric spectroscopy and machine learning methods for extra virgin olive oils discrimination based on the olive cultivar and ripening stage, *Food Measure*, Vol. 17, 2023, pp. 2917–2931, https://doi.org/10.1007/s11694-023-01836-5.

[24] X. Wu, S. Gao, Y. Niu, Z. Zhao, R. Ma, B. Xu, H. Liu, and Y. Zhang, Quantitative analysis of blended corn-olive oil based on Raman spectroscopy and one-dimensional convolutional neural network, *Food Chemistry*, Vol. 385, 2022, pp. 1–6, https://doi.org/10.1016/j.foodchem.2022.132655.

[25] H. Zhao, Y. Zhan, Z. Xu, J. J. Nduwamungu, Y. Zhou, R. Powers, and C. Xu, The application of machine-learning and Raman spectroscopy for the rapid detection of edible oils type and adulteration, *Food Chemistry*, Vol. 373, 2022, pp. 1–9, https://doi.org/10.1016/j.foodchem.2021.131471.

[26] B. Ozen, C. Cavdaroglu, and F. Tokatli, Trends in authentication of edible oils using vibrational spectroscopic techniques, *Analytical Methods*, Vol. 26, 2024, pp. 4216–4233, https://doi.org/10.1039/D4AY00562G.

[27] F. H. Baltork, S. V. Zade, Y. Mazaheri, A. M. Alizadeh, H. Rastegar, Z. Adbian, M. Torbati, and S. A. Damirchi, Recent methods in detection of olive oil adulteration: State-of-the-Art, *Journal of Agriculture and Food Research*, Vol. 16, 2024, pp. 1–16, https://doi.org/10.1016/j.jafr.2024.101123.

[28] W. Song, N. Jiang, H. Wang, and J. Vincent, Use of smartphone videos and pattern recognition for food authentication, *Sensors and Actuators B: Chemical*, Vol. 304, 2020, pp. 1–8, https://doi.org/10.1016/j.snb.2019.127247.

[29] E. Kahya, and Y. Aslan, Comparative Analysis of Deep Learning Models for Olive Detection on the Branch, *WSEAS Transactions on Computers*, Vol. 22, 2023,

pp. 338–351, https://doi.org/10.37394/23205.2023.22.39.

[30] N. H. Bahtiti, F. M. A. Orabi, M. H. Kailani, I. A. Rahman, A. Nahle, Z. O. Alfaouri, and H. H. A/ Abdallat, A Comparative LC/MS Analysis of Jordanian Olive Stone, Fruits, Leaves, and Oils, *WSEAS Transactions on Environment and Development*, Vol. 19, 2023, pp. 903–916, https://doi.org/10.37394/232015.2023.19.86.

[31] The LS-55 and LS-45 Fluorescence Spectrofluorometers, Perkin Elmer, [Online]. https://resources.perkinelmer.com/lab-solutions/resources/docs/BRO_LS-55andLS-45FluorescenceSpectrophotometer.pdf (Accessed Date: November 11, 2024).

[32] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Fourth Edition, 2017, https://doi.org/10.1016/C2015-0-02071-8.

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA Data Mining Software: An Update, *ACM SIGKDD Explorations Newsletter*, Vol. 11, No. 1, 2009, https://doi.org/10.1145/1656274.1656278.

[34] Aksion adjective: that which is of superior quality and is worthy of admiration, *Aksion Olives*, [Online]. https://www.aksion.gr/ (Accessed Date: December 27, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.