International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

# Assessing statistical neural networks potentiality to distinguish PDO Kalamata and Molaoi olive oil varieties

Theodoros Anagnostopoulos [1, 2, *], Chara Kottara [2] and Ioakeim Spiliopoulos [1]

[1] *Department of Food Science and Technology, University of Peloponnese, 24100, Kalamata, Greece.*

[2] *Department of Business Administration, University of West Attica, 12241, Athens, Greece.*

International Journal of Science and Research Archive, 2024, 13(01), 2939–2949

Publication history: Received on 08 September 2024; revised on 19 October 2024; accepted on 21 October 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1997

## Abstract

Several regional areas in Greece produce high quality olive oil by cultivating certain varieties. Olive oil varieties of Kalamata and Molaoi are of special interest, since they produce extra virgin olive oil. Concretely, Kalamata is a city located in southwestern Greece. Intuitively, Molaoi is a town located in southeastern Greece. Subsequently, both geographic locations are known for their famous olive oil quality. Continually, Protected Designation of Origin (PDO) Kalamata olive oil, established by Council regulation (EC) No 510/2006, is considered an exceptional extra virgin olive oil variety. Specifically, there is a need to distinguish PDO Kalamata olive oil from other olive oil varieties is Greece such as the Molaoi olive oil, since PDO Kalamata olive oil is the main variety exported in the global olive oil market. Distinguishing PDO Kalamata from Molaoi olive oil is possible by incorporating statistical neural networks. Concretely, applying neural network experimentation enables differentiation between variations of certain chemical characteristics observed in certain geographic locations of Greece. In this paper, we use statistical neural networks to distinguish the geographical origin of PDO Kalamata olive oil compared with Molaoi olive oil based on synchronous excitation–emission fluorescence spectroscopy of provided olive oils samples evaluated in the chemical laboratory. Evaluation based on certain experimentation phase and subsequent data visualization of the adopted statistical neural networks are promising for distinguishing the samples of PDO Kalamata olive oil with high values of prediction accuracy. Such ability enables olive oil industry to assess extra virgin olive oil profitable potentiality in global market.

**Keywords:** PDO Kalamata olive oil; Molaoi olive oil; Synchronous emission-excitation; Fluorescence spectroscopy; Statistical neural networks; Data visualization

## 1. Introduction

In smart farming cropping of plants useful for citizens is a fundamental issue for nutrition [1]. Intuitively, smart agriculture is a significant dimension of the smart city concept, which aims to define methods of efficient geographic cultivation in rural region areas [2]. Concretely, olive oil farmers in the Messenia geographic region of Greece produce the protected designation of origin (PDO) extra virgin olive oil, established by Council regulation (EC) No 510/2006, with the name PDO Kalamata olive oil in the agricultural areas of the Kalamata city. PDO Kalamata olive oil variety is extensively cultivated and produces the extra virgin olive oil with organoleptic properties [3], [4]. Extensive, exploitation of region areas provides farmers the feasibility to gain more income since specific olive oil microclimates affect the quality of the selected variety [5]. To protect olive oil high quality and prevent its adulteration, global governmental agencies like the European Commission, International Olive Council, and Codex Alimentarius have developed standards to regulate olive oil by establishing a set of physical, chemical, and organoleptic characteristics, [6]. Well defined traditional chemical methods incorporated to ensure olive oil quality are focused on the identification and quantification of pre-defined compounds or classes of compounds of olive oil according to the regulations of the

* Corresponding author: Theodoros Anagnostopoulos

above-mentioned global governmental agencies. Such methods are time-consuming as well as demand expensive apparatus. Subsequently, it holds the same for the detection of olive oil adulteration although these methods fail to detect the adulteration from certain adulterants. Concretely, the non-targeted analysis has attracted much research and scientific attention. Proposed approach focuses on screening the olive oil without any prior knowledge of chemical composition. Such research, uses analytical techniques that produce a signal which is affected by all the compounds (i.e., metabolites) present in olive oil. Adopted methods shorten the analysis process but a vast number of data sources are required to perform data analytics based on statistical data analytics models, [7]. Intuitively, to assess the quality of gathered olive oil there is a need to incorporate specific Internet of Things (IoT) devices, such as dedicated sensors and actuators [8]. Subsequently, a device that is commonly used for such a process is fluorescence spectroscopy, which is calibrated accordingly to perform differences of excitation and emission radiation to the olive oil sample [9]. Continually, fluorescence spectrometry has been used extensively in the past years due to its efficient precision in recognizing chemical components of olive oil samples thus exploiting its overall quality [10]. Incorporated technology can access input data from olive oil sample sources to measure optimally the chemical ingredients of a given olive oil sample as well as to be able to discriminate the olive oil quality categories as well as its origin [11]. Concretely, fluorescence spectra technology can detect in high effectiveness adulteration of olive oil with other lower-quality oils, such as sunflower oil or soybean oil [12]. Specifically, collecting samples from different geographical origins of Greece agricultural regions enables the generation of different data sources [13]. Intuitively, exploited data can be visualized and analyzed in deep detail by effective statistical data analytics models. Concretely, the application of statistical classification models enables the classification of olive oil samples into certain categories able to differentiate the quality of each sample [14].

High quality of extra virgin and virgin olive oil have recently attracted consumer interest because of their finest quality, and its potential health benefits derived from their consumption. Intuitively, high price of extra virgin olive oil and its reputation makes olive oil a target for fraudsters. Subsequently, significant research has been performed in the area of olive oils' analysis, classification, authentication, origin, and adulteration. Proposed spectroscopic techniques such as ultraviolet-visible (i.e., UV–Vis) absorption [15], [16], fluorescence spectroscopy, [17], mass spectrometry [18], Raman spectroscopy, [19], nuclear magnetic resonance [20] and FT-NIR [21] have been proposed to classify and detect adulteration and origin of olive oil. Artificial intelligence classification methods based on statistical data analytics is used to compare virgin olive oil quality in, [22]. Concretely, fluorescence spectroscopy is used along with principal component technology and factorial discriminant analysis for monitoring and classifying certain virgin olive oil varieties. Raman spectroscopy is incorporated in, [23], to identify olive oil quality using classification techniques. Intuitively, the adopted method used a one-dimensional convolutional deep-learning neural network to observe optimal classification results. Portable Raman spectroscopy is used in, [24], to provide quality assessment and control of several olive oil varieties. Subsequently, the proposed method adequately covers the cases of adulterated compound low-quality oils within the virgin olive oil. Intuitively, extensive statistical classification and authentication techniques are incorporated in, [25], to distinguish the origins of virgin olive oil. Intuitively, it is proposed an authentication process is proposed to analyze volatile olive oil compounds and chemometrics to assess the quality of certain olive oil varieties within a local geographic area. Statistical data analytics models are incorporated in, [26], to classify specific olive oil varieties. Concretely, the adopted classification method uses discrimination techniques to input machine learning models with spectroscopic data thus achieving effective prediction accuracy of olive oil behavior by exploiting fusion emission and absorption. Subsequently, fluorescence spectroscopy is incorporated in, [27], to classify the high quality of olive oil. Subsequently, the proposed method assesses a certain thermal oxidation technique, which exploits the potentiality of an Ultra Violet (UV) fluorescence spectroscopy system to perform specific imaging classification of extra virgin olive oil varieties.

A machine learning classification model based on time series analyses is incorporated in, [28], which can distinguish several virgin olive oil varieties. Concretely, a statistical transformation of the generated input data sources is performed on each virgin olive oil variety to assess the ensemble classification schema thus observing optimal values of the prediction accuracy evaluation metric. Adulterated olive oil, in [29], can be discriminated with the incorporation of Attenuated Total Reflection (ATR) and FTIR spectroscopy technologies. Concretely, the proposed methods are capable of distinguishing pure samples of virgin olive oil from different oil blends by exploiting the potentiality of partial least squares discriminant analysis (PLS-DA) applied to given olive oil compounds. A multivariate classification analysis is incorporated in, [30], which can distinguish extra virgin olive oils. Concretely, the adopted method is based on Fourier Transform Infrared Spectroscopy (FTIR) along with multivariate analysis to classify accurately virgin olive oils' geographic origins, which come from several producing countries. Specific statistical methods and applications for distinguishing several extra virgin olive oils' local geographic origins are proposed in the literature, [31]. Concretely, the classification of olive oil geographic origins is based on certain chemometric data sources. Subsequently, such chemometric data are generated from several olive oils compounds, which input the fluorescence spectroscopy decision-making models to achieve optimal prediction accuracy. Intuitively, synchronous scanning of chemometric data sources produced by significantly detailed fluorescence spectroscopy measurements is also supported in certain research efforts, [32]. Such detailed knowledge is then exploited by specific statistical classification learning models, which can distinguish several varieties of edible extra virgin olive oils. Edible olive oils' premium quality is assessed in the literature, [33]. Concretely, such ability is achieved by the incorporation of synchronous fluorescence spectroscopy, which can differentiate the quantification of tocopherols from the input olive oil compounds. Intuitively, agriculture geographic origins of olive oil varieties, [34], are feasible due to the incorporation of chemometric analysis. Continually, such advanced analytical methodology, which is applied to data sources can predict olive oil's registered designation with optimal precision taking into consideration synchronous excitation and emission of fluorescence spectra values. Rapid spectroscopic methods (Vis–NIR and FT-MIR) along with PLS analysis were applied to study thermal stress of virgin olive oils, [35]. Specifically, due to the manipulation of generated data sources to certain statistical learning models, which can evaluate optimally spectroscopic and chemometric technologies. Pattern recognition is also incorporated in extra virgin olive oil varieties classification, [36]. Subsequently, near-infrared spectrometry provides the technical methodology to assess the strengths of screening methods, which are then used to authenticate extra virgin olive oils from near local geographic origins. Shelf-Life olive oil varieties are monitored and then classified to certain geographic origins, [37]. Intuitively, IoT sensors and actuators technology is exploited to enhance fluorescence spectroscopy characteristics thus being able to correctly assess the multiclass classification process, which is based on certain statistical learning models.

Concretely, there are many research approaches that deal with the origins of olive oil based on statistical data analytics models. Promising research efforts incorporate generated data from several chemometric technologies. However, data manipulation requires improvement to distinguish data interconnections, which can provide efficient results. In this research paper, fluorescence spectroscopy is exploited by applying enhanced data preprocessing. Intuitively, such optimized data sources are then used by a statistical data analytics model to perform binary classification to distinguish between PDO Kalamata olive oil compared with Molaoi olive oil variety in cultivated within the Greece agricultural region. Specifically, we input synchronous emission-excitation fluorescence spectra, provided by synchronous photoluminescence spectra IoT-enabled technology device, of PDO Kalamata olive oils as well as Molaoi olive oils. Each category of olive oil composes a unique class. Concretely, since the categories are two the classification problem is characterized as binary classification, where one class represents PDO Kalamata olive oil, while the other represents Molaoi olive oil. Intuitively, PDO Kalamata olive oil samples are is cultivated in the areas of: (1) Aris, (2) Thouria, (3) Verga, (4) Arfara, and (5) Meligalas. Subsequently, Molaoi olive oil samples are cultivated in the areas of: (1) North Molaoi, and (2) South Molaoi. Continually, we input such data sources to a specific statistical neural network model to assess its experimental results in recognizing the difference of the two local cultivations. Proposed statistical neural network is evaluated with certain evaluation methods and metrics to observe an optimal binary classification of input olive oil data samples. The outcome of the research paper is to be able to distinguish PDO Kalamata olive oil from Molaoi olive oil enabling relative industry to exploit extra virgin olive oil profitable potentiality into global market.

The rest of the paper is organized as follows. In Section 2 it is defined the adopted materials and research methods, which are incorporated in the current research effort. Section 3, presents observed results based on certain experimentation process and performs detailed discussion on the derived outputs of the research paper, while Section 4 concludes the paper and proposes future work to the research community.

## 2. Material and methods

### 2.1. Data model

Data sources provided to perform experimental analytics are synchronous emission-excitation fluorescence spectra. Concretely, such spectra were 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) was adjusted to 30 nm [38]. Subsequently, observed spectra in the current research paper were recorded at $\Delta\lambda = 30$. Concretely, the excitation and emission slit were tuned to 4 nm. The scan rate was 50nm/min. Intuitively, each olive oil sample was measured triplicate using the new freshly prepared solution. Intuitively, each measurement of an olive oil sample was statistically handled as a different sample of the same origin. Incorporated data sources are composed of specific structure. Continually, some of the observed data sources are collected from PDO Kalamata olive oil produced in local areas in the rural areas of the city of Kalamata in the Messenia region, while other provided data were collected from Molaoi olive oil produced in local areas of the Laconia region. Since this is a binary classification problem there are observed two classes, namely Class 0 and Class 1. Let us define Class 0 assigned to PDO Kalamata olive oil while Class 1 assigned to Molaoi olive oil. Concretely, data collected from PDO Kalamata olive oil are in total 29 olive oil samples from the local cultivation areas of: (1) Aris, (2) Thouria, (3) Verga, (4) Arfara, and (5) Meligalas. Intuitively, the distribution of collected data samples for PDO Kalamata Class 0 are as follows: (1) 2 samples from Aris, (2) 2 samples from Thouria, (3) 7 samples from Verga,

(4) 15 samples from Arfara, and (5) 3 samples from Meligalas. Intuitively, data collected from Molaoi olive oil are in total 18 olive oil data samples for the local cultivation areas of: (1) North Molaoi, and (2) South Molaoi. Concretely, the distribution of collected data samples for Molaoi olive oil Class 1 are as follows: (1) 7 samples from North Molaoi, and (2) 11 samples from South Molaoi. Concretely, total number of data samples are 47, (i.e., 29 are from Class 0 and 18 are from Class 1)

*2.1.1. Data structure*

Intuitively, it holds that synchronous emission-excitation fluorescence spectra are composed of certain dimensional samples, where the first 5 dimensions denote the 5 predictive attributes, while the last 1 dimension denote the class attribute, such as $(p_i, c_j)$. Concretely, let us define the 5 predictive attributes as $p_i$, where $i \in [1, 5]$ is the identifier of each predictive attribute, where: $i = 1$ refers to tocopherols, $i = 2$ refers to phenolic compounds, $i = 3$ refers to oxidation products of triglycerides, $i = 4$ refers to oxidation products of tocopherols, and $i = 5$ refers to chlorophyll's predictive attributes. Intuitively, let us define the class attribute as $c_j$, where $j \in [0,1]$ is the identifier of the class attribute value, where: $j = 0$ refers to Class 0 value denoting PDO Kalamata olive oil, while $j = 1$ refers to Class 1 value denoting Molaoi olive oil.

*2.1.2. Data visualization*

Concretely, there is a need to visualize provided data classes to observe in detail their structure. The distribution of the adopted classes (i.e., Class 0 for PDO Kalamata olive oil and Class 1 for Molaoi olive oi)l should be plotted to be able to distinguish the different nature of the examined class attributes. Fig. 1 visualizes the two classes, where it can be observed that in the lower left side of Fig. 1 there are presented the 29 data samples assigned to Class 0. Specifically, PDO Kalamata olive oil sample instances could be distinguished since they are denoted with small blue 'x' marks. Concretely, it can be also observed that in the upper right side of Fig. 1 there are depicted the 18 data samples assigned to Class 1. Subsequently, Molaoi olive oil sample instances could be distinguished since they are plotted as small red 'x' marks.

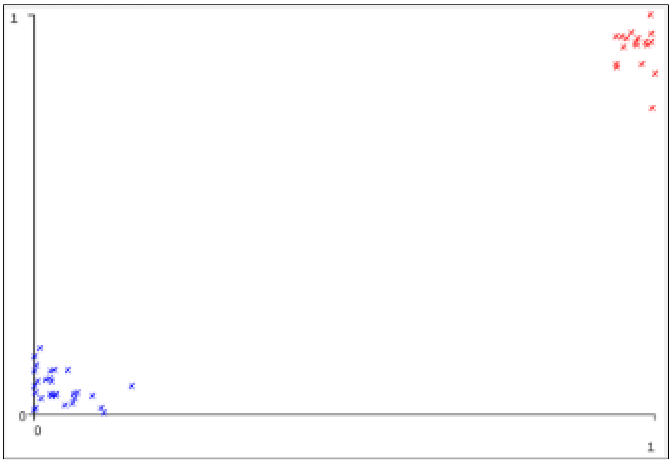


**Figure 1** Data classes visualization (i.e., Class 0 of PDO Kalamata olive oil and Class 1 of Molaoi olive oil)

## 2.2. Evaluation parameters

A deep understanding of the results requires to incorporate certain parameters able to assess the performance of the adopted statistical neural network model. Intuitively, evaluation methods and evaluation metrics should be incorporated to perform specific experiments and observe derived results.

*2.2.1. Evaluation method*

Evaluation of the statistical neural network model requires certain evaluation methods. In this research effort authors adopt one of the widely used evaluation methods, due to its simplicity and optimum results, which is 10-fold cross-validation, [39]. Specifically, such an 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. Concretely, this process is repeated until all the parts are used for training and testing. The proposed evaluation method is adopted in the data analytics methodology since it provides effective results based on certain input data able to explain the observed data source's predictive analytics behavior.

*2.2.2. Evaluation metrics*

Based on the adopted evaluation method, which is proposed to support the experimental setup there is a need to incorporate specific evaluation metrics. Specifically, there are adopted several metrics, which are: (1) prediction accuracy, (2) correctly classified instances, and (3) confusion matrix that can assess the effectiveness of a statistical neural network model.

*2.2.3. Prediction accuracy*

Effectiveness of the adopted statistical neural network model is assessed by incorporating prediction accuracy evaluation metric, $a \in [0, 1]$, which is defined in the following mathematical equation, (1):

$$a = \frac{tr_{pos}+tr_{neg}}{tr_{pos}+fl_{pos}+tr_{neg}+fl_{neg}} \text{……….}(1)$$

It holds that, $tr_{pos}$, are the instances, which are classified correct as positives, and, $tr_{neg}$, are the instances, which are classified correct as negatives. In addition, $fl_{pos}$, are the instances, which are classified false are positives, and, $fl_{neg}$, are the instances, that are classified false as negatives. A low value of $a$ means a weak classifier while a high value of a indicates an efficient statistical neural network classifier. 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 results.

*2.2.4. Correctly classified instances*

Concretely, in statistical data analytics, it is common to express prediction accuracy as a percentage thus observed results being more easily interpreted and presented. Concretely, it is used the term correctly classified instances, $c \in [0\%, 100\%]$, which is defined according to the following mathematical equation, (2):

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

Where, a value close to 0% means that the classification model is not efficient, while a value close to 100% indicates that the statistical neural network model is able to classify instances optimally.

*2.2.5. Confusion matrix*

Intuitively, adopted statistical neural network classification 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 two classes, (i.e., Class 0: PDO Kalamata olive oil, and Class 1: Molaoi olive oil) has the following encoded form, as described in Table 1. It holds that, "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, while "D" quantity depicts the number of Class 1 instances, which are correctly classified as instances of Class 1. A given classification model is considered efficient if it maximizes the elements of the main diagonal of the confusion matrix (i.e., "A" and "D") and minimizes the other elements. A confusion matrix is incorporated in data analytics evaluation methodology to support effectiveness and explain in deep detail the statistical nature of output experimental results observed by the prediction accuracy evaluation metric

## 3. Results

### 3.1. Experiments and results

Proposed data model, which is pre-processed according to two class values (i.e., Class 0: PDO Kalamata olive oil and Class 1: Molaoi olive oil) is incorporated to perform specific experiments and observe derived results. It holds that an experimental setup is necessary to formulate the experimental phase with certain evaluation methods and metrics and observe the results of the current research paper.

*3.1.1. Experimental setup*

Concretely, certain parameters are incorporated to set up the experimental process. Intuitively, it is defined the number of classes (i.e., Class 0 and Class 1), which is assigned to each data sample instance. Subsequently, predictive attributes used to describe a certain class attribute, which are defined accordingly. Continually, a specific statistical neural network model is adopted to perform the experiments and observe the output results.

*3.1.1.1. Binary classification*

Due to the fact that the number of class values is two this classification process is characterized as a binary classification problem. Concretely, two classes are defined as follows: (1) Class 0: PDO Kalamata olive oil, and (2) Class 1: Molaoi olive oil. Subsequently, the number of predictive attributes is five, which are characterized 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'. The number of data sample instances is 47, where 29 samples are assigned to Class 0 while the remaining 18 are assigned to Class 1. Specifically, samples distribution for Class 0 is the following: (1) 2 samples from Aris, (2) 2 samples from Thouria, (3) 7 samples from Verga, (4) 15 samples from Arfara, and (5) 3 samples from Meligalas. Intuitively, samples distribution for Class 1 is as follows: (1) 7 samples from North Molaoi, and (2) 11 samples from South Molaoi.

**Table 1** Binary classification confusion matrix

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

**Table 2** Spectra values ranges of Class 0

| **Predictive attributes** | **Min** | **Max** |
|---|---|---|
| Tocopherols | 131.867 | 136.868 |
| Phenolic compounds | 14.247 | 19.246 |
| Oxidation products of triglycerides | 3.859 | 8.358 |
| Oxidation products of tocopherols | 0.004 | 3.495 |
| Chlorophylls | 52.481 | 56.982 |

Specific numerical spectra values are recorded by 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) was adjusted to 30 nm for the predictive attributes of Class 0 (i.e., PDO Kalamata olive oil) are observed in the following ranges: (1) tocopherols spectra values are within interval $[131.867, 136.868]$, (2) phenolic compounds values are within range $[14.247, 19.246]$, (3) oxidation products of triglycerides spectra values are within interval $[3.859, 8.358]$, (4) oxidation products of tocopherols values are within range $[0.004, 3.495]$, and (5) chlorophylls spectra values are within interval $[52.481, 56.982]$. Numerical spectra values of Class 0 (i.e., PDO Kalamata olive oil) are presented in Table 2.

**Table 3** Spectra values ranges of Class 1

| **Predictive attributes** | **Min** | **Max** |
|---|---|---|
| Tocopherols | 166.936 | 171.937 |
| Phenolic compounds | 3.035 | 7.036 |
| Oxidation products of triglycerides | 0.092 | 2.657 |
| Oxidation products of tocopherols | 0.092 | 2.657 |
| Chlorophylls | 30.202 | 35.201 |

Concretely, the numerical values of the predictive attributes of Class 1 (i.e., Molaoi olive oil) take values as described in following ranges: (1) tocopherols spectra values are within interval $[166.936, 171.937]$, (2) phenolic compounds values are withing range $[3.035, 7.036]$, (3) oxidation products of triglycerides spectra values are within interval $[0.511, 5.262]$, (4) oxidation products of tocopherols values are within range $[0.092, 2.657]$, and (5) chlorophylls spectra values are within interval $[30.202, 35.201]$. Numerical values of Class 1 (i.e., Molaoi olive oil) are presented in Table 3.

#### *3.1.1.2. Neural network statistical classification model*

Certain experiments were performed to be able to select the optimum statistical neural network model, which is efficient for the examined binary classification problem. Specifically, we performed experiments with several machine learning classifiers available in the Weka machine learning software, [40], [41]. Continually, the machine learning model, which has optimal predictive behavior emerged to be experimentally the neural network statistical data analytics model, which is composed by three layers (i.e., an input layer, a hidden layer and an output layer). Concretely, input layer is composed by 10 nodes, hidden layer is composed by 20 nodes, while output layer is composed by 10 nodes. Since such model achieves optimal performance it is adopted for further experimentation to observe the derived output results of the current research effort.

### *3.1.2. Derived results*

Performing evaluation of the experimental requires to define a specific evaluation method (i.e., 10-fold cross-validation) and metrics used to assess the efficiency of the adopted statistical machine learning model, which in this case is the neural network model. Subsequently, based on certain evaluation parameters specific derived results are observed, which define the effectiveness of the incorporated experimental setup adopted in the current research paper. Intuitively, to understand the observed results and be able to explain the research effort's findings it is significant to use the incorporated evaluation method and evaluation metrics. Such knowledge would reveal the inherent complexity that exists in the provided data sources aiming to observe optimal results for the adopted statistical neural network model.

Observed classification visualization

By applying neural network data analytics model to the supplied data sources for both classes (i.e., Class 0 and Class 1) there are observed promising results denoting a data analytics model with impact during the prediction process. Concretely, in Fig. 2 it can be observed that only two samples of Class 0 are falsely classified as samples of Class 1 annotated with a small blue square at the upper right side of the graph. Subsequently, all the other data samples are correctly classified by the adopted statistical neural network model.


**Figure 2** Classification errors visualization of the Class 0 PDO Kalamata and Class 1 Molaoi olive oil classes.

Observed prediction accuracy

Certain evaluation method, which is incorporated to evaluate the adopted binary classification statistical neural network model is 10-fold cross-validation. According to the adopted evaluation method observed prediction accuracy is: $a = 0.9574$, which is a high value for prediction accuracy thus proving that the adopted statistical neural network model is suitable for the examined binary classification problem. Intuitively, the high value observed for the prediction accuracy enables the adopted neural network data analytics model to be incorporated for similar use in new unseen olive oil instances in a further future research that might extend the potentiality of the current research effort to both examined geographical regions of interest (i.e., Kalamata city and Molaoi town).

Observed correctly classified instances

Based on the adopted evaluation method of 10-fold cross-validation correctly classified instances it occurred to be: $c = 95.74\%$, which indicated that the selected statistical data analytics model is an efficient choice for the examined classification problem.

Observed confusion matrix

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

**Table 4** Observed results of the binary classification

| Class 0 | Class 1 | ← Classified as |
|---|---|---|
| 27 | 2 | Class 0 |
| 0 | 18 | Class 1 |

Intuitively, it can be observed that most 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. Subsequently, such an optimal prediction behavior indicates a classification model with prediction impact for the examined binary classification problem. Intuitively, such a detailed confusion matrix enables the observation of experimental results in detail thus being able to assess the effectiveness of the adopted data analytics neural network model for differentiating PDO Kalamata olive oil in comparison with Molaoi olive oil based on specific experimental instances.

## 4. Discussion

As observed form the definition of the problem it indicates a binary classification problem of two discrete classes (i.e., Class 0: PDO Kalamata olive oil and Class 1: Molaoi olive oil), with five separate numerical predictive attributes and a total of 47 data instances. Specifically, 29 sample instances are assigned to Class 0 while the remaining 18 are assigned to Class 1. Intuitively, current research effort has achieved significant values of the observed results based on certain evaluation metrics, which indicate the stability of the examined evaluation parameters. Concretely, observed prediction accuracy and correctly classified instances have relatively high result values. Subsequently, binary confusion matrix has high values in the main diagonal denoting an efficient and accurate predictive model. Subsequently, adoption of statistical neural network data analytics model for binary classification required in the current research paper results in effective and stable results able to differentiate PDO Kalamata olive oil in comparison with Molaoi olive oil based on the adopted methodological research framework.

## 5. Conclusion

To identifying the differences between the two regions that produce extra virgin olive oil, such PDO Kalamata olive oil and Molaoi olive oil is of highly importance in the current research paper. Specifically, the incorporated synchronous photoluminescence spectra of olive oils IoT device can provide initial data sources to compare the origins of PDO Kalamata olive oil and Molaoi olive oil. In this research paper, we use statistical data analytics models to perform binary classification between Class0 (i.e., PDO Kalamata olive oil) and Class 1 (i.e., Molaoi olive oil) collected from several geographic origins. Observed evaluation of the statistical neural network model are based on certain methods and metrics, which have proved to be promising for distinguishing PDO Kalamata olive oil compared with Molaoi olive oil. Concretely, according to the research outcomes, future work should mainly focus on the incorporation of more detailed input measurement data sources based on improvements in synchronous photoluminescence spectra IoT-enabled technology, thus providing a more robust input to the selected statistical neural network classification model. Intuitively, current research paper could be further used in deep detail to verify authentication and to detect adulteration of PDO Kalamata olive oil thus facing the fraud problem occurring in the olive oil global market currently.

## Compliance with ethical standards

### *Acknowledgments*

Authors would like to acknowledge the Department of Food Science and Technology of the University of Peloponnese in Kalamata, Greece for providing the experimental olive oil input data samples from PDO Kalamata, Messenia and Molaoi, Laconia geographic regions, which incorporated in the current research study to observe the presented output results.

*Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Boer J. L. D., Erickson B. Setting the Record Straight on Precision, Agriculture Adoption. Agronomy Journal. 2019; 111(4): 1552-1569.

[2] Paiho S., Tuominen P., Rokman J., Ylikerala M., Pajula J., Siikavirta H. Opportunities of collected city data for smart cities. IET Smart Cities. 2022; 4(4): 275-291.

[3] Trabelsi L., Ncube B., Hassena A. B., Zouairi M., Amar F. B., Gargouri K. Comparative study of productive performance of two olive oil cultivars Chemlali Sfax and Koroneiki under arid conditions. South African Journal of Botany. 2023; 154: 356-364.

[4] Miao X., Ma J., Miu X., Zhang H., Geng Y., Hu W., Deng Y., Li N. 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. 2023; 288: 1-12.

[5] Issa A., Riachy M. E., Mitri C. B., Doumit J., Skaff W., Karam L. Influence of geographical origin, harvesting time and processing system on the characteristics of olive-mill wastewater: A step toward reducing the environmental impact of the olive oil sector. Environmental Technology & Innovation. 2023; 32: 1-12.

[6] Aparicio R., Morales M. T., Ruiz R. A., Tena N., González D. L. G. Authenticity of olive oil: Mapping and comparing official methods and promising alternatives. Food Research International. 2013; 54(2): 2025-2038.

[7] Ellis D. I., Muhamadali H., Haughey S. A., Elliott C. T., Goodacre R. 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. 2015; 7(22): 9375-9716.

[8] Rajak P., Ganguly A., Adhikary S., Bhattacharya S. Internet of Things and smart sensors in agriculture: Scopes and challenges. Journal of Agriculture and Food Research. 2023; 14: 1-13.

[9] Krause J., Gruger H., Gebauer L., Zheng X., Knobbe J., Pgner T., Kicherer A., Gruna R., Langle T., Beyerer J. Smart Spectrometer–Embedded Optical Spectroscopy for Applications in Agriculture and Industry. Sensors. 2021; 21: 1-18.

[10] Karoui R., Blecker C. Fluorescence Spectroscopy Measurement for Quality Assessment of Food Systems – a Review. Food Bioprocess Technology. 2011; 4: 364-386.

[11] Khani S., Ghasemi J. B., Vanak Z. P. Development of computer vision system for classification of olive oil samples with different harvesting years and estimation of chlorophyll and carotenoid contents: A comparison of the proposed method's efficiency with UV-Vis spectroscopy. Journal of Food Composition and Analysis. 2024; 106078: 21-42.

[12] Drakopoulou S. K., Kritikou A. S., Baessmann C., Thomaidis N. Untargeted 4D-metabolomics using Trapped Ion Mobility combined with LC-HRMS in extra virgin olive oil adulteration study with lower-quality olive oils. Food Chemistry. 2024; 434: 1-9.

[13] Schiano M. E., Sodano F., Cassiano C., Magli E., Seccia S., Rimoli M. G, Albrizio S. Monitoring of seven pesticide residues by LC-MS/MS in extra virgin olive oil samples and risk assessment for consumers. Food Chemistry. 2024; 442: 1-8.

[14] Reda R., Saffaj T., Bouzida I., Saidi O., Belgrir M., Lakssir B., Hadrami E. M. E. Optimized variable selection and machine learning models for olive oil quality assessment using portable near infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2023; 303: 1-11.

[15] Santos R. A., Cancilla J. C., Pérez A. P., Moral A., Torrecilla J. S. Quantifying binary and ternary mixtures of monovarietal extra virgin olive oils with UV–vis absorption and chemometrics. Sensors and Actuators B: Chemical. 2016; 234: 115-121.

[16] Milanez K. D. T. M., Nóbrega T. C. A., Nascimento D. S., Insausti M., Band B. S. F., Pontes M. J. C. Multivariate modeling for detecting adulteration of extra virgin olive oil with soybean oil using fluorescence and UV–Vis spectroscopies: A preliminary approach. LWT – Food Science and Technology. 2017; 85: 9-15.

[17] Merás I. D., Manzano J. D., Rodríguez D. A., Peña A. M. Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitation–emission profiles combined with multi–way classification. Talanta. 2018; 178: 751-762.

[18] Girolamo F. D., Masotti A., Lante I., Scapaticci M., Calvano C. D., Zambonin C., Muraca M., Putignani L. A. Simple and effective mass spectrometric approach to identify the adulteration of the mediterranean diet component extra–virgin olive oil with corn oil. International Journal of Molecular Sciences. 2015; 16: 20896-20912.

[19] Li Y., Fang T., Zhu S., Huang F., Chen Z., Wang Y. Detection of olive oil adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and SiPLS. Spectrochimica Acta Part A: Molecular Biomolecular Spectroscopy. 2018; 189: 37-43.

[20] Rotondo A., Mannina L., Salvo A. Multiple Assignment Recovered Analysis (MARA) NMR for a Direct Food Labeling: The Case Study of Olive Oils. Food Analytical Methods. 2019; 12: 1238-1245.

[21] Mossoba M. M., Azizian H., Kia A. R. F., Karunathilaka S. R., Kramer J. K. G. First Application of Newly Developed FT–NIR Spectroscopic Methodology to Predict Authenticity of Extra Virgin Olive Oil Retail Products in the USA. Lipids. 2017; 52: 443-455.

[22] Zaroual H., Chene C., Hadrami E. M. E., Karoui R. Comparison of four classification statistical methods for characterizing virgin olive oil quality storage up to 18 months. Food Chemistry. 2022; 370: 1-16.

[23] Wu X., Gao S., Niu Y., Zhao Z., Xu B., Ma R., Liu H., Zhang Y. Identification of olive oil in vegetable blend oil by one-dimensional convolutional neural network combined with Raman spectroscopy. Journal of Food Composition and Analysis. 2022; 108: 1-7.

[24] Barros I. H. A. S., Paixao L. S., Nascimento M. H. C., Lacerda V. J., Figueiras P. R., Romao W. Use of portable Raman spectroscopy in the quality control of extra virgin olive oil and adulterated compound oils. Vibrational Spectroscopy. 2021; 116: 1-10.

[25] Cecchi L., Migliorini M., Giambanelli E., Rosseti A., Cane A., Mulinacci N., Melani F. Authentication of the geographical origin of virgin olive oils from the main worldwide producing countries: A new combination of HS-SPME-GC-MS analysis of volatile compounds and chemometrics applied to 1217 samples. Food Control. 2020; 112: 1-10.

[26] Stefas D., Gyftokostas N., Kourelias P., Nanou E., Kokkinos V., Bouras C., Couris S. Discrimination of olive oils based on the olive cultivar origin by machine learning employing the fusion of emission and absorption spectroscopic data. Food Control. 2021; 130: 1-8.

[27] Rotich V., Riza D. F. A., Giametta F., Suzuki T., Ogawa Y., Kondo N. Thermal oxidation assessment of Italian extra virgin olive oil using an UltraViolet (UV) induced fluorescence imaging system. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2020; 237: 1-8.

[28] Bagnall A., Davis L., Hills J., Lines J. Transformation Based Ensembles for Time Series Classification. Proceedings of the 2012 SIAM International Conference on Data Mining (SDM). Anaheim. California. USA. 2012; April 26-28: 307-318.

[29] Mata P. D. L., Vidal A. D., Sendra J. M. B., Medina A. R., Rodriguez L. C., Canada M. J. A. Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. Food Control. 2012; 23: 449-455.

[30] Tapp H. S., Defernez M., Kemsley E. K. FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origins of Extra Virgin Olive Oils. Journal of Agricultural and Food Chemistry. 2003; 51: 6110-6115.

[31] Sikorska E., Khmelinskii I., Sikorski M. Analysis of Olive Oils by Fluorescence Spectroscopy: Methods and Applications. InTech. 2012; 4: pp. 63 – 88.

[32] Sikorska E., Gorecki T., Khmelinskii I. V., Sikorski M., Koziol J. Classification of edible oils using synchronous scanning fluorescence spectroscopy. Food Chemistry. 2005; 89: 217-225.

[33] Sikorska E., Swiglo A. G., Khmelinskii I., Sirorski M. Synchronous Fluorescence of Edible Vegetable Oils. Quantification of Tocopherols. Journal of Agriculture and Food Chemistry. 2005; 53(18): 6988-6994.

[34] Dupuy N., Dreau Y. L., Ollivier D., Artaud J., Pinatel C., Kister J. Origin of French Virgin Olive Oil Registered Designation of Origins Predicted by Chemometric Analysis of Synchronous Excitation-Emission Fluorescence Spectra. Journal of Agricultural and Food Chemistry. 2005; 53(24): 9361-9368.

[35] Maggio R. M., Valli E., Bendini A., Caravaca A. M. G., Toschi T. G., Cerretani L. A spectroscopic and chemometric study of virgin olive oils subjected to thermal stress. Food Chemistry. 2011; 127: 216-221.

[36] Bertran E., Blance M., Coello J., Iturriaga H., Maspoch S., Montoliu I. Near infrared spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins. Journal of Near Infrared Spectroscopy. 2000; 8: 45-52.

[37] Prieto A. L., Tena N., Ruiz R. A., Gonzalez D. L. G., Sikorska E. Monitoring Virgin Olive Oil Shelf-Life by Fluorescence Spectroscopy and Sensory Characteristics: A Multidimensional Study Carried Out under Simulated Market Conditions. Foods. 2020; 9: 1-20.

[38] Vokhmintsev A. S., Minin M. G., Henaish A. M. A., Weinstein I. A. Spectrally resolved thermoluminescence measurements in fluorescence spectrometer. Measurement. 2015; 66: 90-94.

[39] Trevor H., Tibshirani R., Friedman J. H. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York. Springer. 2nd Edition. 2009.

[40] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. The WEKA Data Mining Software: An Update. SIGKDD Explorations Newsletter. 2009; 11(1): 10-18.

[41] Frank E., Hall M. A., Witten I. H. The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 4th Edition. 2016.