Extraction of physicochemical properties from the fluorescence spectrum with 1D convolutional neural networks: Application to olive oil

Journal of Food Engineering 336 (2023) 111198

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

Extraction of physicochemical properties from the fluorescence spectrum with 1D convolutional neural networks: Application to olive oil

Francesca Venturini [a,c,*], Michela Sperti [b], Umberto Michelucci [c], Arnaud Gucciardi [c,d], Vanessa M. Martos [e], Marco A. Deriu [b]

[a] Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401, Winterthur, Switzerland
[b] PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy
[c] TOELT Llc, Machine Learning Research and Development, Birchlenstr. 25, 8600, Dübendorf, Switzerland
[d] Artificial Intelligence Laboratory, University of Ljubljana, Ljubljana, Slovenia
[e] Department of Plant Physiology, Faculty of Sciences, Biotechnology Institute, Campus Fuentenueva S/n, 18071, University of Granada, Spain

ARTICLE INFO

Keywords:
Convolutional neural networks
Artificial intelligence
Machine learning
Fluorescence spectroscopy
Optical sensor
Olive oil
Quality control 2010 MSC
00–01
99–00

ABSTRACT

One of the main challenges for olive oil producers is the ability to assess oil quality regularly during the production cycle. The quality of olive oil is evaluated through a series of parameters that can be determined, up to now, only through multiple chemical analysis techniques. This requires samples to be sent to approved laboratories, making the quality control an expensive, time-consuming process, that cannot be performed regularly and cannot guarantee the quality of oil up to the point it reaches the consumer. This work presents a new approach that is fast and based on low-cost instrumentation, and which can be easily performed in the field. The proposed method is based on fluorescence spectroscopy and one-dimensional convolutional neural networks and allows to predict five chemical quality indicators of olive oil (acidity, peroxide value, UV spectroscopic parameters K_{270} and K_{232}, and ethyl esters) from one single fluorescence spectrum obtained with a very fast measurement from a low-cost portable fluorescence sensor. The results indicate that the proposed approach gives exceptional results for quality determination through the extraction of the relevant physicochemical parameters. This would make the continuous quality control of olive oil during and after the entire production cycle a reality.

1. Introduction

Determining the quality of olive oil is an expensive and complex procedure that requires a chemical analysis by specialized laboratories and organoleptic evaluation by accredited testing panels. For producers it is thus impossible to determine olive oil quality effectively and frequently enough during the production process. Olive oil quality assessment is important as the chemical composition changes dramatically with time (Gó mez-Coca et al., 2016) depending on, for example, storage and temperature conditions. Both the chemical parameters and the procedures for their determination (methods ranging from titration to gas chromatography) are specified in the European regulation (Commission regulation, 1991) and amendment (Commission implementing regulation no 1348, 2013). These regulations provide a decision tree for the verification if an olive oil class is consistent with the declared quality.

The challenge of determining olive oil quality is fundamental, as olive oil plays an important role in the cultural and culinary heritage of the Mediterranean countries, and its demand has grown in the latest years to other regions of the world. The growing interest, particularly in its highest quality grade, extra virgin olive oil (EVOO), is due to its high nutritional value, its richness in bioactive molecules (Serrano et al., 2021), and its importance to our health due to its content of anti-inflammatory and antioxidant substances. For these reasons, extra virgin olive oil (EVOO) is a fundamental ingredient of the dietary pattern known worldwide as the “Mediterranean diet”, which has been associated with important health benefits, such as the reduction of the prevalence of cardiovascular and metabolic diseases (Uylaşer and Yildiz, 2014; Fabiani, 2016; Gorzynik-Debicka et al., 2018).

Fluorescence spectroscopy has attracted significant research efforts in the last years, as it offers a rapid, cost-efficient, and at the same time sensitive technique to investigate the properties of vegetable oils (Karoui and Blecker, 2011; Kongbonga et al., 2011; Sikorska et al., 2012; Al Riza et al., 2021). Several fluorescent compounds are naturally present in olive oil, like pigments such as chlorophyll and beta-carotene, phenolic

* Corresponding author. Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401 Winterthur, Switzerland.
E-mail address: vent@zhaw.ch (F. Venturini).

https://doi.org/10.1016/j.jfoodeng.2022.111198
Received 17 April 2022; Received in revised form 6 June 2022; Accepted 28 June 2022
Available online 2 July 2022

compounds, such as tocopherol, and primary and secondary oxidation products (Mart í n-Tornero et al., 2021). These compounds are related to the quality criteria established in the European regulation. It is therefore of great importance to develop methods for extracting those physicochemical information from fluorescence spectra.

The extraction of information from the spectral data can be a difficult task depending on the type of data acquired, which may range from a single spectrum to more complex excitation emission matrices (EEMs), synchronous scanning data (Skoog et al., 2017) or near-infrared spectroscopy (Yuan et al., 2020). Typical approaches consist in multivariate analysis techniques and classification methods, like for example, principal component analysis (PCA), partial least squares regression (PLS), and PLS discriminant analysis (PLS-DA) to mention only a few.

The use of artificial neural networks is known to be a useful tool, particularly because it does not require a preprocessing of the data or a dimensionality reduction (Michelucci, 2018). Several reviews describe the application of statistical and machine learning methods, including neural networks, to the analysis and quality determination of olive oil (Sikorska et al., 2014; Zaroual et al., 2021; Meenu et al., 2019; Gonzalez-Fernandez et al., 2019; Aroca-Santos et al., 2019; Lastra-Mejias et al., 2019). Feed-forward neural networks have been up to now successfully employed for classification purposes starting from fluorescence data (Venturini et al., 2021), but do not offer sufficient flexibility for more complex tasks that analyse data that have some kind of spatial structure (like two-dimensional images or one-dimensional optical spectra). To address this issue, various architectures, such as vision transformers or convolutional neural networks, have been applied to the classification problem of vegetable oils (Zhao et al., 2022) with fluorescence data.

One-dimensional neural networks (1D-CNNs) are more efficient architectures when dealing with one-dimensional input data, as recent works have shown for spectroscopic classification (Acquarelli et al., 2017), electrocardiography real-time classification (Kiranyaz et al., 2015), for chemometric analysis from, for example, near-infrared reflectance spectra, and near- and mid-infrared absorption spectra (Malek et al., 2018).

By drastically reducing the requirements on the measuring hardware and on the quality of data, this work presents a novel method to extract the physicochemical properties relevant for the quality characterization of virgin olive oil from fluorescence spectra using 1D-CNN. The spectra can be acquired with a very simple and compact sensor from undiluted and unprepared samples. To the best of our knowledge, this is the first time that all key parameters are extracted simultaneously, without pre- and post-processing of the data from a simple fluorescence spectrum. The limitations and further development possibilities are discussed in the conclusions.

The contributions of this paper are four. Firstly, it describes an approach based on 1D convolutional neural networks to extract the five physicochemical characteristics relevant for the determination of olive oil's quality from one single fluorescence spectrum. The method is described with guidelines and criteria for the implementation. The application to small datasets with a leave-one-out cross validation technique is discussed in detail. Secondly, this approach does not require a technical training once the neural network has been trained and, therefore, has a high impact and applicability in the olive oil industry. Thirdly, by using a sensor based on low-cost components this approach enables a democratisation of olive oil quality control. Finally, the method is demonstrated by the application on a dataset of Spanish oils and shows, for the first time, that it is possible to compete for quantitative analysis with complex chemical analysis, for example, chromatography, using a simple and fast optical measuring method supported by convolutional neural networks in one dimension.

2. Materials and methods

2.1. Olive oil samples

In this study, 22 virgin olive oils of three qualities were investigated: extra virgin olive oil (EVOO), virgin olive oil (VOO), and lampante olive oil (LOO). For the definition of the quality classes the reader is referred to (Commission regulation, 1991). The oils were provided by the producer Conde de Benalúa, Granada, southern Spain, from the 2019–2020 harvest. All samples were analyzed by accredited laboratories for chemical and organoleptic properties according to the current European regulation (Commission regulation, 1991; Commission implementing regulation no 1348, 2013). The selected properties relevant to this study are listed in Table 1.

Fig. 1 shows the decision tree for the determination the quality of the olive oil described in the European regulation (Commission regulation, 1991) and its amendments (Commission implementing regulation no 1348, 2013) and illustrates the list of parameters investigated in this study. The same parameters are investigated in this study. Note the parameter ΔK is not considered in this study since the differences between the measured values for almost all oils were within the experimental error reported by the accredited laboratories. In supervised learning *DeltaK* takes the role of the target variables that the model has to learn. Since the differences of the values are of the same order of magnitude of the reported experimental error, the model will learn the inherent noise present in the instrumental error.

2.2. Instrumentation

The fluorescence spectra were taken with a portable sensor specifically designed to have a simple construction and a compact design (Fig. 2). The light is provided by an excitation UV LED, that can be exchanged. In this study, three wavelengths were investigated: 340 nm, 365 nm, and 395 nm. These excitation wavelengths were chosen because they correspond to maxima in the absorption band of the fluorophores present in olive oil, such as chlorophylls (Ferreiro-Gonz á lez et al., 2017; Torreblanca-Zanca et al., 2019; Borello and Domenici, 2019). The oil samples were placed into commercial transparent 4 ml glass vials, taking care that no headspace was present to reduce oxidation. The fluorescence is collected by a miniature spectrometer (STS-Vis,

Table 1
List of the olive oils samples analyzed in this study including selected physicochemical characteristics. EVOO: extra virgin olive oil, VOO: virgin olive oil, LOO: lampante olive oil.

Label	Acidity (%)	Peroxide value (mEq O_2/kg)	K_{270}	K_{232}	Ethyl esters (mg/Kg)	Quality
D03	0.35	8.4	0.123	1.435	26	VOO
D04	0.34	8.6	0.108	1.403	40	VOO
D05	0.36	10.3	0.112	1.44	18	VOO
D06	0.31	9.2	0.151	1.484	18	VOO
D07	0.50	8.9	0.150	1.537	47	VOO
D08	0.40	8.5	0.158	1.546	25	VOO
D19	0.25	4.9	0.13	1.540	10	EVOO
D20	0.26	4.6	0.14	1.540	10	EVOO
D35	0.17	6.4	0.12	1.63	8	EVOO
D38	0.16	6.4	0.12	1.63	9	EVOO
D45	0.17	4.9	0.12	1.63	7	EVOO
D46	0.18	5.0	0.13	1.63	8	EVOO
D47	0.18	5.2	0.13	1.64	16	EVOO
D49	0.9	9.9	–	–	–	LOO
D51	2.16	–	–	–	–	LOO
D52	1.78	22	–	–	–	LOO
D53	0.7	8.7	–	–	–	LOO
D64	0.2	7.1	0.13	1.63	29	VOO
D73	0.2	8.9	0.14	1.66	15	EVOO
D77	0.24	10.4	0.13	1.74	26	VOO
D81	0.16	4.9	0.12	1.63	9	EVOO
D92	0.18	5	0.17	1.91	15	EVOO

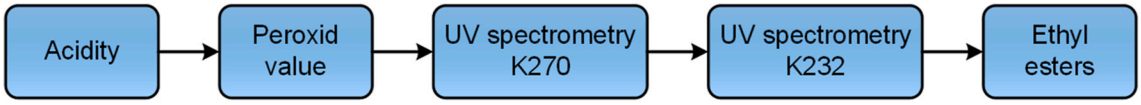

Fig. 1. Sequence of parameters to be analyzed for the verification of olive oil quality. Adapted from (Commission regulation, 1991; Commission implementing regulation no 1348, 2013).

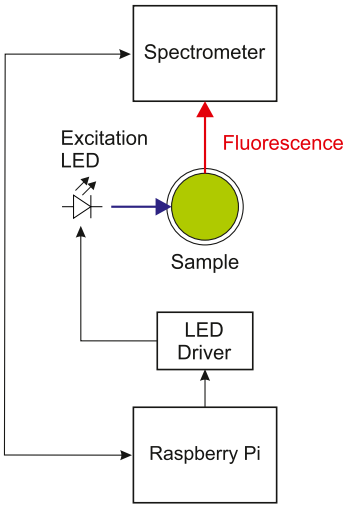

Fig. 2. Schematics of the portable fluorescence sensor. Blue: excitation light, red: fluorescence light.

Ocean Optics, USA) placed at 90° with respect to the LED to avoid the excitation light transmitted by the sample to reach the spectrometer. Both the LED driver and the spectrometer are controlled by a Raspberry Pi. The details of the sensor are reported in (Venturini et al., 2021).

All measurements in this work were performed on undiluted samples. Although fluorescence in olive oil is subjected to the inner filter effect (Skoog et al., 2017), the problem is not relevant for the analysis discussed in this work. In fact, the fluorescence is intense enough that the strong absorption does not decrease the signal-to-noise ratio, and possible sample-dependent effects are learned and compensated by the artificial neural network model. For each olive oil sample, 20 spectra were taken, each acquired with 1 s integration time. All spectra were acquired under identical conditions (illumination intensity, integration time, and geometry) to be able to quantitatively compare the different intensities.

2.3. Dataset preparation

Since the oils were measured by different laboratories and are of different qualities, the amount of data available per oil varies. For example, for some LOO oils like D49 or D52, only the acidity and peroxide value were measured. If the value of the parameter is missing, such a sample was not considered for the model predicting such parameter. Therefore, the number of oils available for the estimation of the chemical parameters depends on the parameter itself. The number of samples considered for each parameter is listed in Table 2.

Table 2
Number of olive oils samples used for the training and test of the CNN for each parameter.

Parameter	Number of samples
Acidity	22
Peroxide value	21
K_{270}	18
K_{232}	18
Ethyl esters	18

All spectra are normalized after the dark background is subtracted so that each of the spectra has an average of 0 and a standard deviation of 1.

2.4. Convolutional neural network model

The model developed in this work is shown in Fig. 3 and consists of a one-dimensional convolutional neural network (1D-CNN) with one convolutional layer, followed by a max-pooling and a second convolutional layer with finally two dense layers and an output layer with one single neuron with the identity activation function. The interested reader can refer to (Michelucci, 2019) for a mathematical description of CNNs. This choice was inspired by previous studies, where 1D-CNNs with two or three convolutional layers were applied to different spectroscopic data, for example, reflectance spectra and Raman spectra (Malek et al., 2018; Zhang et al., 2019; Liu et al., 2018). The idea behind the sequence of layers is that the first layer extracts rough data patterns, and the subsequent layers learn more high-level abstractions. A convolutional layer is characterized by the number of filters and their sizes. During the 1D convolution operation, each filter is convolved across the length of the input array, computing the dot product between the filter entries and the input, producing a one-dimensional array (called feature map) for each of the filters (Michelucci, 2019).

In a CNN the learnable parameters are the filters themselves that are learned by backpropagation (Michelucci, 2019; LeCun et al., 1989; Gu et al., 2018).

The parameters varied and tested in this work were the number of filters in the first convolutional layer (4 and 6), the number of filters in the second convolutional layer (4 and 6), the pooling size (8 and 16), the number of epochs (5000, 10,000) and the mini-batch size (8, 16, and 64). As activation function for all the layers (except the output one) the ReLU function was chosen.

The size of the filters and their number were chosen based on the fluorescence spectra characteristics and instrument properties. Previous studies suggest that the number of expected features contained in the fluorescence spectra of olive oils is limited. Possible examples are the height of the main fluorescence peak, its width, the area under the peak, and the area under the second fluorescence peak (Torreblanca-Zanca et al., 2019; El Orche et al., 2020). For this reason, the number of filters to test was chosen to be 4 and 6. Additionally, since the spectrometer resolution is ca. 30 pixels, the size of the filters was chosen to be 40. This reflects the fact that spectral features with a bandwidth smaller than the resolution of the spectrometer are convolved with the instrument response function. Choosing a size of 40 pixels for the filters prevents the network from considering much too granular information that the spectrometer cannot extract due to its resolution, with the additional positive effect that overfitting will be reduced. The layers are designed to perform feature extraction, and indirectly a dimensionality reduction, so to extract a very low number of features, by doing first max pooling and then a second convolution operation with filters of half the size of those in the first convolutional layer. At the end, two small dense layers have the task to perform the regression to finally extract the chemical parameter selected. The CNN was implemented using the TensorFlow ™ Python library. All the models were trained with backpropagation (Kelley, 1960) from scratch. No pre-trained model was used.

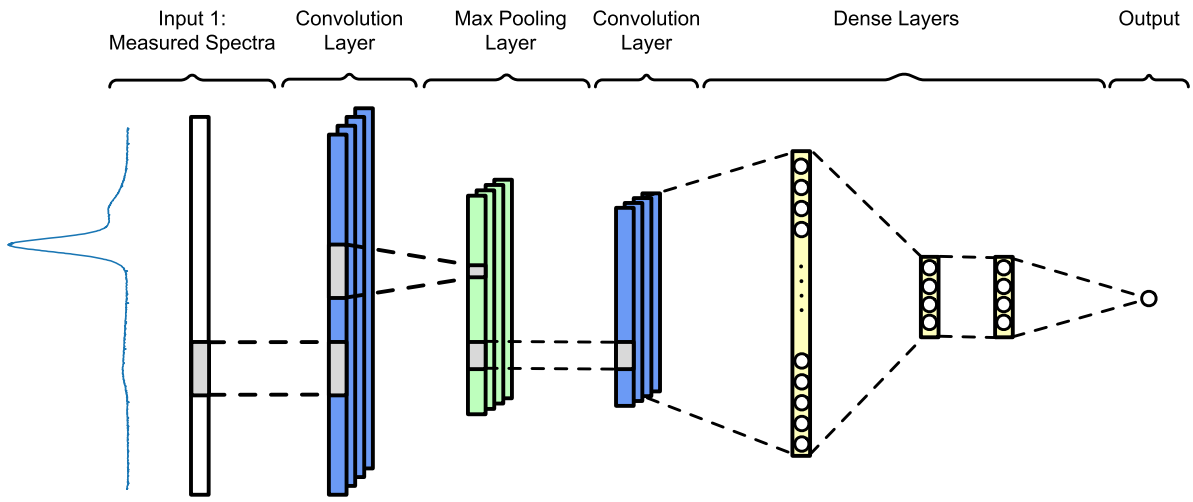

Fig. 3. A schematic representation of the 1D-CNN used in this paper. The blue layers are convolutional ones, the green max pooling layers and the yellow marked ones are dense layers. The output layer has 1 neuron with the identity activation function.

2.5. Metrics, performance evaluation and validation

The metrics used to evaluate the model performance are two: the mean squared error (MSE) and the mean absolute error (MAE). The MSE was used as loss function for the training of the neural networks (Michelucci, 2018), while the MAE was used to determine the prediction performance of the neural network. Indicating the expected (true) value of the parameters for the ith spectrum and the predicted value from the neural network with y_i and $\hat{y}_i$ respectively, the two metrics can be expressed with the following formulas:

$$\begin{aligned} \mathrm{MSE} &= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \\ \mathrm{MAE} &= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} |\hat{y}_i - y_i| \end{aligned} \tag{1}$$

where $N = 20 \cdot N_{oil}$ is the number of spectra composing the dataset (N is the product of 20 repetitions for each of the N_{oil} oils measured). Since the dataset is small, a leave-one-out cross-validation approach (Michelucci and Venturini, 2021) was used to determine the generalization properties of the network. In such an approach the (20) spectra of one single oil are removed from the dataset and used for validation, while the network is trained on the spectra of all remaining oils. This procedure is repeated for each oil, therefore resulting in N values of the metrics evaluated for all the oils. The results reported in this paper are thus the average $\langle \mathrm{MAE} \rangle$ and standard deviation $\sigma(\mathrm{MAE})$ of N values. A risk of the leave-one-out cross-validation is that the neural network may simply learn to predict the value of the parameter corresponding to the oil left out for all the oils. Therefore, it is quite important to always check training predictions to make sure that $\langle \mathrm{MAE} \rangle$ evaluated on the training and validation dataset are comparable. For each of the N_{oil} in the leave-one-out cross-validation two models during training were saved: the one with the lowest value of the loss function evaluated on the validation set (with the left-out oil), and the one with the lowest value of the loss function on the training set (with $N_{oil} - 1$). The model that showed comparable values for $\langle \mathrm{MAE} \rangle$ for training and validation dataset was then chosen.

To choose which set of hyper-parameters (number and size of filters, pooling size, epochs, etc.), normally one would select the network parameters that give the lowest value of the chosen metric (in this case $\langle \mathrm{MAE} \rangle$ on the validation dataset). However, this approach cannot be used directly here, as there is some variability (measured by the variance of the MAE) in the results and many of the calculated averages overlap within one standard deviation. Therefore, it is important to determine if the different models in the hyper-parameter-tuning phase give results that are statistically different. This can be checked with a t-test described in detail in A. The results showed that changing the number of filters and their size gives results that are not significantly different, therefore by using Occam's razor decision criteria (Hiroshi, 2022) the simplest network was chosen for the results presented in this paper. The chosen network has 6 filters of size 40 in the first convolutional layer, and 4 filters with size 20 in the second convolutional layer. A decreasing number of filters in the first and second layers (6 and then 4 respectively) was chosen to facilitate a progressive and more stable feature extraction process (Michelucci, 2018). Finally, a pooling size of 8 and a dropout rate of 0.5 were taken.

10,000 epochs produced better results than 5000 consistently, therefore the former value was chosen. The mini-batch size was chosen to be 64 when saving the best model on the validation dataset, and 16 when saving the best model on the training dataset.

3. Results and discussion

3.1. Fluorescence spectra of olive oil

The fluorescence signals measured with the portable sensor at 340 nm were extremely weak and are hardly detectable with the sensor used in this study. For this reason, they are not reported here. The raw fluorescence spectra of all the oils obtained with excitation at 365 nm and at 395 nm are shown in Fig. 4. For clarity, the spectra are shown divided into the three quality classes EVOO, VOO, and LOO. Each curve of Fig. 4 shows one single spectrum after background subtraction, without averaging or smoothing.

The fluorescence spectrum of all oils is characterized by a strong intensity in the region between 650 nm and 750 nm, with an intense peak at circa 678 nm and a weaker broader one at ca. 722 nm, typical of chlorophyll and pheophytins (Hern á ndez-S á nchez et al., 2017; Mishra et al., 2018; Baltazar et al., 2020; Galeano D í az et al., 2003). The strongest peak, however, shows variations in the spectra position and intensity towards higher wavelengths, which are particularly significant in LOOs. These variations are consistent with previous results (Torreblanca-Zanca et al., 2019). The spectra obtained with excitation at 365 nm and 395 nm are similar, with slightly higher fluorescence intensities for 395 nm

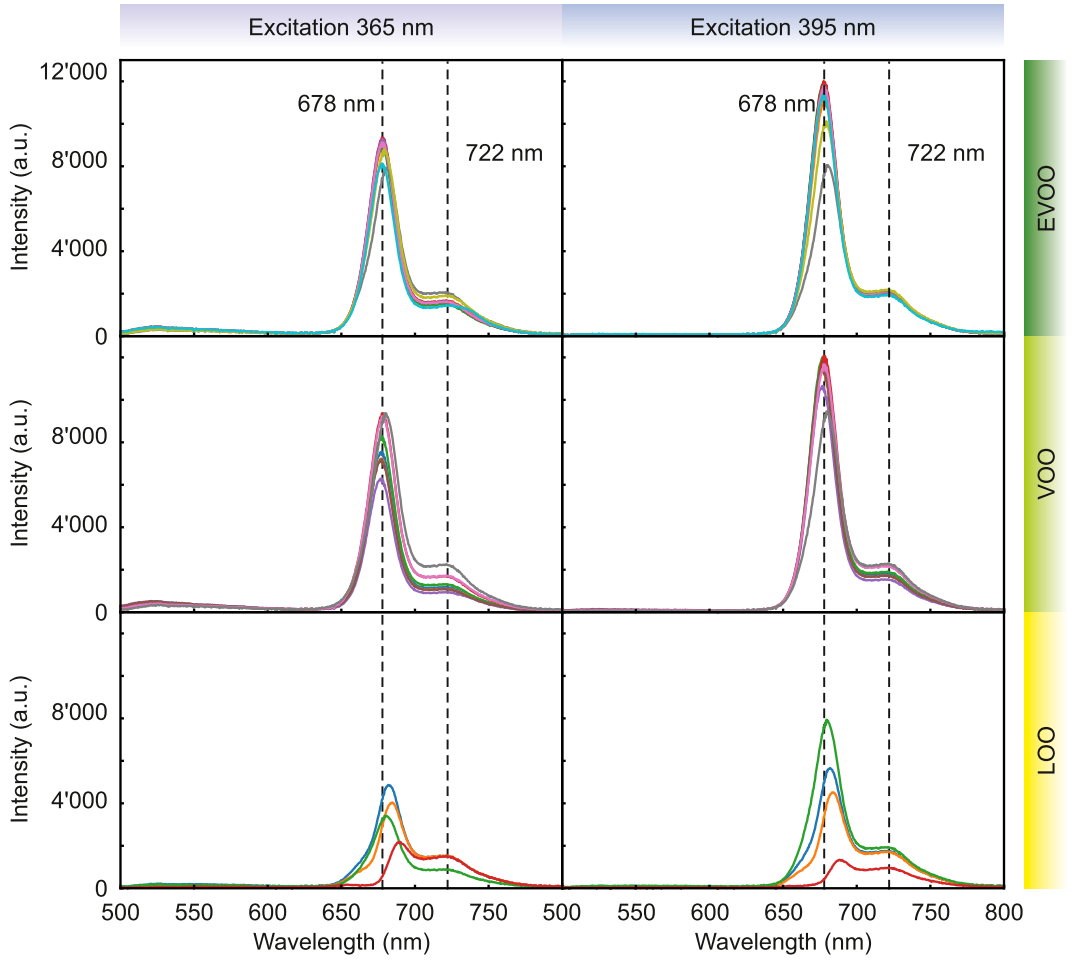

Fig. 4. Fluorescence emission spectra of the measured olive oils divided in the quality classes EVOO, VOO and LOO. On the left: spectra obtained with excitation at 365 nm; on the right: spectra obtained with excitation at 395 nm. Each curve shows a single spectrum without averaging or smoothing.

excitation. This is consistent with the stronger absorption expected around 400 nm (Torreblanca-Zanca et al., 2019; Borello and Domenici, 2019). Noticeably, the fluorescence intensity below 650 nm is present only in the spectra obtained with excitation at 365 nm and is characterized by a weaker absorption peak at ca. 525 nm, attributed to vitamin E (Kyriakidis and Skarkalis, 2000).

3.2. Artificial neural networks results

The 1D-CNN described in Sec. 2.4 was trained to learn to predict each parameter of the decision tree (Fig. 1) successively. The performance is illustrated by plotting the predicted values for each oil against the expected (measured) value, labelled here as true values. The results are illustrated in Fig. 5. The grey area in each panel marks the uncertainty on the true values due to the experimental error, calculated as the average of the error reported by the accredited laboratory on the measured value.

Fig. 5 panel A) shows that the 1D-CNN can predict the acidity exceptionally well, except for two LOO, D51 and D52, which have values well above the 0.8% limit for EVOO. This can be easily understood due to the lack of samples from which the neural networks can learn for acidity values above 1%: since the cross-validation is performed with a leave-one-out method, the 1D-CNN has only one single oil to learn from for acidity values above 1%.

Fig. 5 panel B) shows the results for the prediction of the peroxide value. Also in this case the 1D-CNN can predict the value of the parameter exceptionally well. With exception of the LOO D52 and two other oils, all the predictions are within the average measurement error.

In panel C) and D) of Fig. 5 the predictions for the two UV-spectroscopy parameters K_{270} and K_{232} are shown. For these two parameters, the experimental error is large compared to the range of the values of the parameter. Also here, the predictions remain well within the grey area, showing that the 1D-CNN can learn to predict both UV-spectroscopy parameters within the experimental error.

Finally, panel E) shows the performance for the prediction of the ethyl esters. Here, the 1D-CNN correctly predicts several oils but has more difficulties in the prediction of others. The authors attribute part of the problem to the limited number of oils, but also to the uncertainty of the labels. Differently from the other parameters, the ethyl esters measured by the accredited laboratories were reported with errors ranging from ± 2 to ± 8 mg/kg. Additionally, for the 1D-CNN to learn from the spectra, the parameter must possess a direct or indirect physicochemical signature in the fluorescence. Due to the simplicity of the sensor of this study, the fluorescence signature may be insufficiently strong or clear. Nevertheless, the method described here can give a fast and inexpensive qualitative indication of the ethyl esters without the use of gas chromatography.

The analysis at 365 nm is similar to the one performed at 395 nm, suggesting that similar information is contained in the spectra.

The results can be quantified by calculating the metric $\langle \text{MAE} \rangle$ and its standard deviation $\sigma(\text{MAE})$, evaluated with leave-one-out cross-validation on both the training and the validation dataset. The results for all parameters are reported in Table 3. In the table are also listed the average relative error, calculated as the MAE divided by the true label for each oil and then averaged over all the oils, and the relative label error, calculated as the experimental error divided by the measured

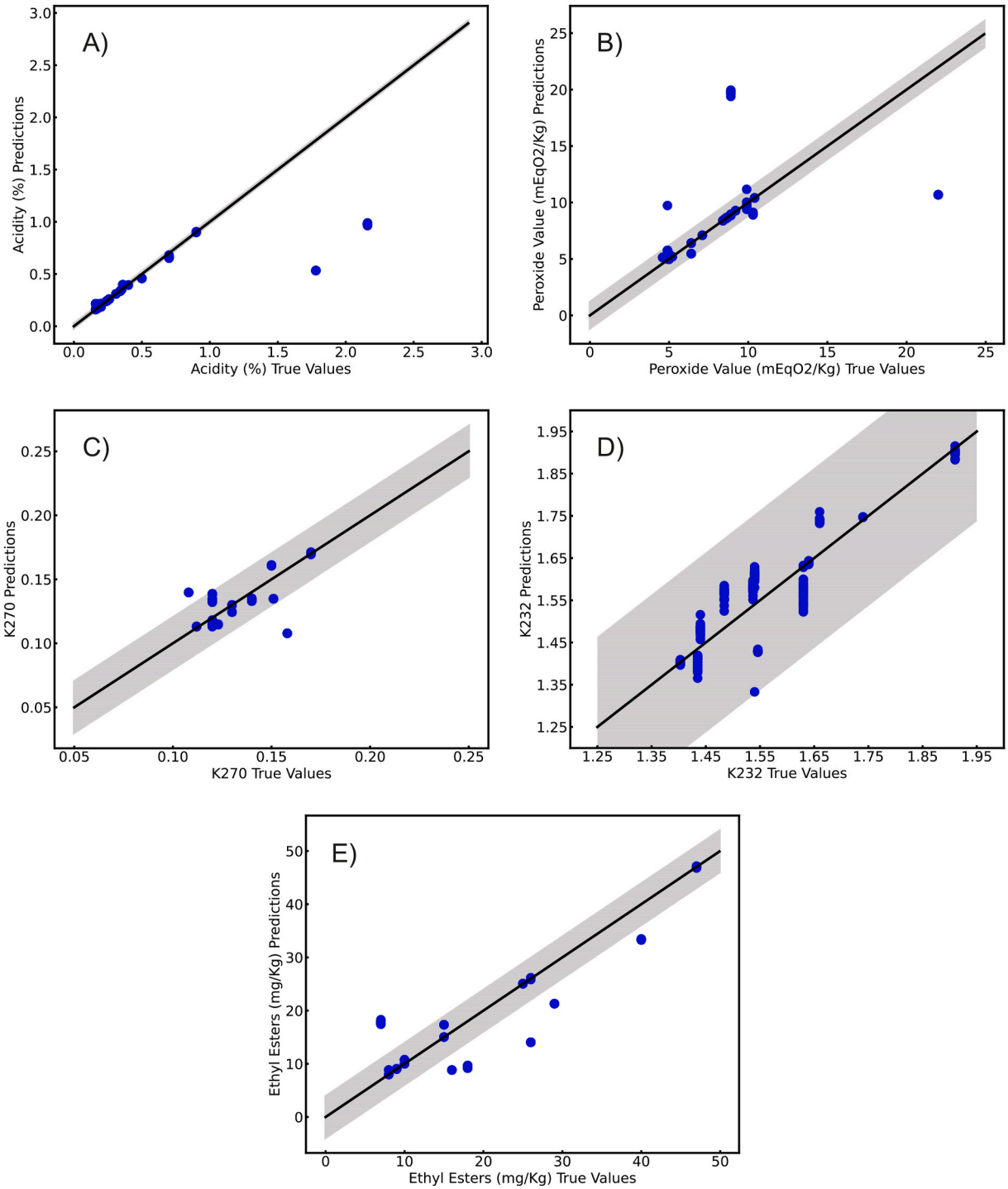

Fig. 5. Comparison of the predicted and true values for all the parameters. Panel A) acidity, panel B) peroxide value, panel C) K_{270}, panel D) K_{232} and panel E) ethyl esters. The solid line corresponds to predictions equal to the true labels. The grey area illustrates the experimental error on the true values. The yellow area marks the range of acceptability for EVOO.

value (true label) for each oil and then averaged over all the oils.

The similarity of the ⟨MAE⟩ for the training and for the validation datasets shows that the 1D-CNN learns effectively and that the models are robust and do not incur the risks associated with the leave-one-out cross-validation. Furthermore, Table 3 shows that the average error is always lower (for the parameters peroxide value and ethyl esters) or much lower (for the parameters K_{270} and K_{232}) than the experimental error from the measurements of accredited laboratories. Only for the acidity, the average error in the prediction is slightly higher than the label error but this is due to the lack of LOO oils with acidity higher than 1% from which to learn, ad discussed before.

The results of Table 3 demonstrate that the proposed approach allows the quantitative assessment of all parameters relevant for the quality control of olive oil. The data for the assessment are acquired with a portable, low-cost compact sensor, without any sample preparation, and required neither pre- nor post-processing. Therefore, the entire quality assessment procedure is fast and can easily be performed in the field, without special training or instrumentation.

Table 3
Summary of the results for all parameters. ⟨MAE⟩ is the mean of the MAE, and σ(MAE) its standard deviation. *T*: Training, *V*: Validation. Average error indicates the average relative error in % of the 1D-CNN prediction, label error indicates the relative experimental error in % on the true label.

Parameters	$\langle MAE_T \rangle$	$\sigma(MAE_T)$	$\langle MAE_V \rangle$	$\sigma(MAE_V)$	Average error (%)	Label error (%)
Acidity (%)	0.10	0.05	0.12	0.35	10	8
Peroxide Value (mEqO2/Kg)	1.01	0.65	1.31	3.19	12	17
K_{270}	0.008	0.003	0.010	0.013	7	15
K_{232}	0.03	0.02	0.04	0.04	2.5	13
Ethyl Esters (mg/Kg)	3.1	1.6	3.6	4.3	23	28

The limitations to the performance observed are due to the limited number of oils available in this study and to the distribution of the values of the parameters (most notably seen for the acidity, where only two oils have values in the upper range). On the other hand, it must be noted that the single origin of the olive oil samples, and thus their similar chemical characteristics, makes the task of extraction of chemical parameters easier. For a more heterogeneous dataset of olive oils it is expected that a more complex architecture will be necessary, as well as a larger dataset.

4. Conclusions

The results in this paper show clearly how the proposed method could substitute the multiple chemical analyses currently needed to assess the quality of olive oil and thus helps producers to keep the quality of their oils under continuous control. The 1D-CNN used in this work was designed to account for the sensor characteristics (e.g., resolution) and the knowledge of the problem (e.g., expected number of features in the spectrum). As a result, the method has shown a very promising performance: from the simple fluorescence spectra it is possible to predict, within the typical experimental errors, all five physicochemical characteristics necessary for quality assessment of olive oil. Of course, one should note that the dataset size in this study is small and, therefore, the results should only be considered as an indication of the potential of the method. Naturally, a larger dataset would allow a more complete analysis of the generalisation properties of such models. Nonetheless, this method has the advantage of using a portable, low-cost and compact instrument, dost not need any sample handling and no data processing. Therefore, it can be used by the anyone on-site and without any scientific training.

The potential of this approach is very promising. For example, by having multiple samples from multiple years, and using meteorological data of the geographical location of production it could be possible to correlate quality with information such as the amount of precipitation, temperature, and so on. This would pave the road to predicting quality based on external factors, probably one of the greatest challenges in the olive oil economy.

As briefly mentioned one of the challenges to be solved in the future is the application of this approach to olive oil samples coming from different producers, different geographical locations, or from harvests of different years. It is to be expected that the chemical signatures in the fluorescence spectra will not be similar anymore between those subgroups, making the prediction of the parameters a much greater challenge. In this case, more complex 1D-CNN architectures and larger datasets will be necessary to keep into account the heterogeneity in the olive oil samples.

Finally, this approach is not limited to olive oil but can be extended to other substances, making the results described here a very promising indication of what could be achieved through one-dimensional convolutional neural networks applied to optical spectra.

Funding

This work was supported by the projects: “VIRTUOUS” funded by the European Union’s Horizon 2020 Project H2020-MSCA-RISE-2019 Grant No. 872181; “SUSTAINABLE” funded by the European Union’s Horizon 2020 Project H2020-MSCA-RISE-2020 Grant No. 101007702; “Project of Excellence” from Junta de Andalucia-FEDER-Fondo de Desarrollo Europeo 2018. Ref. P18–H0-4700.

Credit author statement

Francesca Venturini: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing; **Michela Sperti**: Data curation, Software; **Umberto Michelucci**: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing; **Arnaud Gucciardi**: Data curation, Software; **Vanessa M. Martos**: Funding acquisition, Resources; **Marco A. Deriu**: Funding acquisition, Writing – review & editing.

A. Statistical testing of equivalence of averages

Given two sets of hyper-parameters, indicated here with the subscripts 1 and 2, one can test the equality of the two means of the MAE, $\langle MAE_1 \rangle$ and $\langle MAE_2 \rangle$ respectively, by using the *t*-statistic (Hogg et al., 1977). The formulas used in this paper are based on the ones for confidence intervals for the difference of the means when the variances are unknown and the sample size is relatively small. Note that the *t*-statistics technically works when one deals with normal distributions. In general, the MAE values from the leave-one-out cross-validation approaches have an unknown distribution. However, since one is considering the average, thanks to the central limit theorem, one can assume that the distribution of ⟨MAE⟩ is approximated by a normal distribution (at least one that is not too skewed) and therefore the choice of this approach is justified (Hogg et al., 1977). N_{oil} is of the order of 20 (see Table 1), a number typically considered not large enough for the central limit theorem. Nevertheless, being close to the suggested value of 30, it should give a useful estimate of the statistical significance of the average difference between different sets of hyperparameters. The null-hypothesis H_0 that the two means are equal is rejected if the observed value of

$$T = \frac{\langle MAE_1 \rangle - \langle MAE_2 \rangle}{S_P \sqrt{2/N_{oil}}} \tag{A.1}$$

where

$$S_p = \left[\frac{(N_{oil}-1)\mathrm{Var}(\mathrm{MAE}_1)}{2N_{oil}-2} + \frac{(N_{oil}-1)\mathrm{Var}(\mathrm{MAE}_2)}{2N_{oil}-2}\right]^{1/2} \quad \text{(A.2)}$$

is larger than $t_\alpha(2N_{oil}-2)$ (Hogg et al., 1977) (right-trail probability of size α for the t-distribution with $2N_{oil}-2$ degrees of freedom, or in other words the value that satisfy that the probability $P(t \geq t_\alpha) = \alpha$) for some chosen value of α. For this work, $\alpha = 0.05$ was chosen.

References

Acquarelli, J., van Laarhoven, T., Gerretzen, J., Tran, T.N., Buydens, L.M., Marchiori, E., 2017. Convolutional neural networks for vibrational spectroscopic data analysis. Anal. Chim. Acta 954, 22–31.

Al Riza, D.F., Kondo, N., Rotich, V.K., Perone, C., Giametta, F., 2021. Cultivar and geographical origin authentication of Italian extra virgin olive oil using front-face fluorescence spectroscopy and chemometrics. Food Control 121, 107604.

Aroca-Santos, R., Lastra-Mejías, M., Cancilla, J.C., Torrecilla, J.S., 2019. Linear and non-linear quantification of extra virgin olive oil, soybean oil, and sweet almond oil in blends to assess their commercial labels. J. Food Compos. Anal. 75, 70–74.

Baltazar, P., Hernández-Sánchez, N., Diezma, B., Lleó, L., 2020. Development of rapid extra virgin olive oil quality assessment procedures based on spectroscopic techniques. Agronomy 10 (1), 41.

Borello, E., Domenici, V., 2019. Determination of pigments in virgin and extra-virgin olive oils: a comparison between two near uv-vis spectroscopic techniques. Foods 8 (1), 18.

Commission implementing regulation no 1348/2013 of december 17 2013. Off. J. Eur. Union 338, 2013, 31–67.

Commission regulation (eec) no. 2568/91 of 11 july 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis official journal l 248, 5 september 1991. Offic. JL 248, 1991, 1–83.

El Orche, A., Bouatia, M., Mbarki, M., 2020. Rapid analytical method to characterize the freshness of olive oils using fluorescence spectroscopy and chemometric algorithms. J. Anal. Method. Chem. 2020, 1–9.

Fabiani, R., 2016. Anti-cancer properties of olive oil secoiridoid phenols: a systematic review of in vivo studies. Food Funct. 7 (10), 4145–4159.

Ferreiro-González, M., Barbero, G.F., Álvarez, J.A., Ruiz, A., Palma, M., Ayuso, J., 2017. Authentication of virgin olive oil by a novel curve resolution approach combined with visible spectroscopy. Food Chem. 220, 331–336.

Gómez-Coca, R.B., Fernandes, G.D., del Carmen Pérez-Camino, M., Moreda, W., 2016. Fatty acid ethyl esters (faee) in extra virgin olive oil: a case study of a quality parameter. LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 66, 378–383. https://doi.org/10.1016/j.lwt.2015.10.063. https://www.sciencedirect.com/science/article/pii/S0023643815302814.

Galeano Díaz, T., Durán Merás, I., Correa, C.A., Roldán, B., Rodríguez Cáceres, M.I., 2003. Simultaneous fluorometric determination of chlorophylls a and b and pheophytins a and b in olive oil by partial least-squares calibration. J. Agric. Food Chem. 51 (24), 6934–6940.

Gonzalez-Fernandez, I., Iglesias-Otero, M., Esteki, M., Moldes, O., Mejuto, J., Simal-Gandara, J., 2019. A critical review on the use of artificial neural networks in olive oil production, characterization and authentication. Crit. Rev. Food Sci. Nutr. 59 (12), 1913–1926.

Gorzynik-Debicka, M., Przychodzen, P., Cappello, F., Kuban-Jankowska, A., Marino Gammazza, A., Knap, N., Wozniak, M., Gorska-Ponikowska, M., 2018. Potential health benefits of olive oil and plant polyphenols. Int. J. Mol. Sci. 19 (3), 686.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., 2018. Recent advances in convolutional neural networks. Pattern Recogn. 77, 354–377.

Hernández-Sánchez, N., Lleó, L., Ammari, F., Cuadrado, T.R., Roger, J.M., 2017. Fast fluorescence spectroscopy methodology to monitor the evolution of extra virgin olive oils under illumination. Food Bioprocess Technol. 10 (5), 949–961.

S. Hiroshi, last retrieved on 20th March 2022. [link]. URL https://math.ucr.edu/home/baez/physics/General/occam.html.

Hogg, R.V., Tanis, E.A., Zimmerman, D.L., 1977. Probability and Statistical Inference, vol. 993. Macmillan, New York.

Karoui, R., Blecker, C., 2011. Fluorescence spectroscopy measurement for quality assessment of food systems—a review. Food Bioprocess Technol. 4 (3), 364–386.

Kelley, H.J., 1960. Gradient theory of optimal flight paths. ARS J. 30 (10), 947–954.

Kiranyaz, S., Ince, T., Gabbouj, M., 2015. Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 63 (3), 664–675.

Kongbonga, Y.G.M., Ghalila, H., Onana, M.B., Majdi, Y., Lakhdar, Z.B., Mezlini, H., Sevestre-Ghalila, S., 2011. Characterization of vegetable oils by fluorescence spectroscopy. Food Nutr. Sci. 2 (7), 692–699.

Kyriakidis, N.B., Skarkalis, P., 2000. Fluorescence spectra measurement of olive oil and other vegetable oils. J. AOAC Int. 83 (6), 1435–1439.

Lastra-Mejias, M., Izquierdo, M., Torreblanca-Zanca, A., Aroca-Santos, R., Cancilla, J.C., Sepulveda-Diaz, J.E., Torrecilla, J.S., 2019. Cognitive fluorescence sensing to monitor the storage conditions and locate adulterations of extra virgin olive oil. Food Control 103, 48–58.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., 1989. Handwritten digit recognition with a back-propagation network. Adv. Neural Inf. Process. Syst. 2.

Liu, L., Ji, M., Buchroithner, M., 2018. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. Sensors 18 (9), 3169.

Malek, S., Melgani, F., Bazi, Y., 2018. One-dimensional convolutional neural networks for spectroscopic signal regression. J. Chemometr. 32 (5), e2977.

Martín-Tornero, E., Fernández, A., Pérez-Rodriguez, J.M., Durán-Merás, I., Prieto, M.H., Martín-Vertedor, D., 2021. Non-destructive fluorescence spectroscopy as a tool for discriminating between olive oils according to agronomic practices and for assessing quality parameters. Food Anal. Methods 1–13.

Meenu, M., Cai, Q., Xu, B., 2019. A critical review on analytical techniques to detect adulteration of extra virgin olive oil. Trends Food Sci. Technol. 91, 391–408.

Michelucci, U., 2018. Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks. APRESS Media, LLC.

Michelucci, U., 2019. Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection. Springer.

Michelucci, U., Venturini, F., 2021. Estimating neural network's performance with bootstrap: a tutorial. Mach. Learn. Knowl. Extr. 3 (2), 357–373. https://doi.org/10.3390/make3020018. https://www.mdpi.com/2504-4990/3/2/18.

Mishra, P., Lleó, L., Cuadrado, T., Ruiz-Altisent, M., Hernández-Sánchez, N., 2018. Monitoring oxidation changes in commercial extra virgin olive oils with fluorescence spectroscopy-based prototype. Eur. Food Res. Technol. 244 (3), 565–575.

Serrano, A., De la Rosa, R., Sánchez-Ortiz, A., Cano, J., Pérez, A.G., Sanz, C., Arias-Calderón, R., Velasco, L., León, L., 2021. Chemical components influencing oxidative stability and sensorial properties of extra virgin olive oil and effect of genotype and location on their expression. LWT (Lebensm.-Wiss. & Technol.) 136, 110257. https://doi.org/10.1016/j.lwt.2020.110257. https://www.sciencedirect.com/science/article/pii/S0023643820312469.

Sikorska, E., Khmelinskii, I., Sikorski, M., 2012. Analysis of Olive Oils by Fluorescence Spectroscopy: Methods and Applications, Olive Oil-Constituents, Quality, Health Properties and Bioconversions, pp. 63–88.

Sikorska, E., Khmelinskii, I., Sikorski, M., 2014. Vibrational and electronic spectroscopy and chemometrics in analysis of edible oils. In: Cruz, R.M.S., Khmelinskii, I., Vieira, M. (Eds.), Methods in Food Analysis, pp. 201–234.

Skoog, D.A., Holler, F.J., Crouch, S.R., 2017. Principles of Instrumental Analysis. Cengage learning.

Torreblanca-Zanca, A., Aroca-Santos, R., Lastra-Mejias, M., Izquierdo, M., Cancilla, J.C., Torrecilla, J.S., 2019. Laser diode induced excitation of pdo extra virgin olive oils for cognitive authentication and fraud detection. Sensor. Actuator. B Chem. 280, 1–9.

Uylaşer, V., Yildiz, G., 2014. The historical development and nutritional importance of olive and olive oil constituted an important part of the mediterranean diet. Crit. Rev. Food Sci. Nutr. 54 (8), 1092–1101.

Venturini, F., Sperti, M., Michelucci, U., Herzig, I., Baumgartner, M., Caballero, J.P., Jimenez, A., Deriu, M.A., 2021. Exploration of Spanish olive oil quality with a miniaturized low-cost fluorescence sensor and machine learning techniques. Foods 10 (5), 1010.

Yuan, Z., Zhang, L., Wang, D., Jiang, J., Harrington, P. de B., Mao, J., Zhang, Q., Li, P., 2020. Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis. LWT (Lebensm.-Wiss. & Technol.) 125, 109247. https://doi.org/10.1016/j.lwt.2020.109247. https://www.sciencedirect.com/science/article/pii/S0023643820302358.

Zaroual, H., Chénè, C., El Hadrami, E.M., Karoui, R., 2021. Application of new emerging techniques in combination with classical methods for the determination of the quality and authenticity of olive oil: a review. Crit. Rev. Food Sci. Nutr. 1–24.

Zhang, X., Lin, T., Xu, J., Luo, X., Ying, Y., 2019. Deepspectra: an end-to-end deep learning approach for quantitative spectral analysis. Anal. Chim. Acta 1058, 48–57.

Zhao, Z., Wu, X., Liu, H., 2022. Vision transformer for quality identification of sesame oil with stereoscopic fluorescence spectrum image. LWT (Lebensm.-Wiss. & Technol.) 158, 113173. https://doi.org/10.1016/j.lwt.2022.113173. https://www.sciencedirect.com/science/article/pii/S0023643822001086.