Chemical analysis of olive oils from fluorescence spectra thanks to one-dimensional convolutional neural networks

Chemical analysis of olive oils from fluorescence spectra thanks to one-dimensional convolutional neural networks

Michela Sperti[a], Umberto Michelucci[b], Francesca Venturini[c,b], Arnaud Gucciardi[b,d], and Marco A. Deriu[a]

[a]PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy
[b]TOELT llc, Machine Learning Research and Development, Birchlenstr. 25, 8600 Dübendorf, Switzerland
[c]Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401 Winterthur, Switzerland
[d]Artificial Intelligence Laboratory, University of Ljubljana, Ljubljana, Slovenia

ABSTRACT

The chemical analysis of food is essential to monitor and guarantee its quality. The determination of the chemical parameters, like the concentration of particular substances, is performed by specialized laboratories and is a time-consuming and costly process. Therefore, alternative methods with easier handling are of great interest. Among these fluorescence spectroscopy offers great opportunities. Fluorescence spectra are one-dimensional arrays of values already successfully employed together with artificial neural networks for classification problems in chemistry, physics, and other fields. However, the extraction of specific quantities from the spectra poses a much harder challenge. This work analyzes and compares the ability of feed-forward neural networks (FFNN) and one-dimensional convolutional neural networks (1D-CNN) to extract relevant features from fluorescence spectra of olive oils. The results indicate that 1D-CNN, contrary to FFNN, successfully predicts the chemical parameters with high accuracy. The great advantages of the proposed method are: 1) the possibility of using optical methods instead of time-consuming chemical ones, like chromatography, 2) the lack of any special sample handling, like dilution and 3) the lack of any pre-processing of the data. The problem of small datasets, which may arise for novel techniques like the proposed one, is also addressed statistically by using the leave-one-out resampling technique.

Keywords: fluorescence spectroscopy; fluorescence sensor; olive oil; machine learning; artificial neural networks; quality control

1. INTRODUCTION

Olive oil is one of the most important ingredients of the daily diet of the population living in Mediterranean countries.[1] Extra virgin olive oil (EVOO) is the highest quality of olive oil, with rich nutritional properties.[2] Due to the increasing demand for EVOO, olive oil quality assessment has become necessary. The European Economic Community (EEC) has defined regulations for the commercial labelling of the different qualities of olive oils.[3] Currently, the evaluation is performed by accredited laboratories for the chemical analysis, and panels for the organoleptic analysis, through a complex, time-consuming, and costly process.

Alternatives to the standard methods to predict the above-mentioned chemical properties more easily and cost-effectively are, therefore, of great interest. Among the approaches, fluorescence spectroscopy is very promising since olive oils contain several natural fluorescence molecules like pigments, phenolic compounds, and their oxidation products. Fluorescence spectra are one-dimensional arrays of values that can be exploited to create a unique fingerprint for different types of olive oils.[4,5] For this reason fluorescence spectroscopy has been successfully applied to the study of the properties of vegetable oils, particularly olive oils.[6,7] The majority of the

Send correspondence to F. Venturini: francesca.venturini@zhaw.ch

previous studies aimed at classifying olive oils into several categories,[8,9] while the quantitative extraction of chemical parameters remains a harder challenge.

New methodologies include the application of artificial neural networks (ANN), which have already been applied to olive oil classification starting from fluorescence spectra.[10,11] ANN offer the advantages that they do not require heavy dataset preprocessing and they can deal with complex and non-linear problems, optimising the network structure according to the task at hand.[12]

This work investigates the different ANN architectures, namely feed-forward neural networks (FFNN) and one-dimensional convolution neural networks (1D-CNN), to determine which is best suited to extract from fluorescence spectra the features relevant to the olive oil quality classification: acidity, peroxide value, K_{270}, K_{232}, and ethyl esters. The results show that 1D-CNN perform better and are able to successfully predict olive oil chemical quantities from fluorescence spectra. The problem of dataset dimension, which is small in this study, is also considered by using the leave-one-out resampling technique. This work paves the way to substitute time-consuming chemical processes with optical methods, without the need of heavy sample handling, or any preprocessing of the data.

2. MATERIAL AND METHODS

The entire workflow of the comparison between the approaches with FFNN and 1D-CNN can be summarised by the following basic steps:

1. Dataset preparation: Generation of features (normalized olive oils raw spectra) and labels (five chemical continuous quantities associated with each spectrum).
2. Feed Forward Neural Networks: Tuning, training, and testing of five different FFNN architectures (one for each chemical parameter to be predicted). Since the chemical quantities are continuous values the learning task was designed as a regression task.
3. 1-D Convolutional Neural Networks: Training and testing of a single simple 1D-CNN architecture (the same for each chemical parameter to be predicted), whose structure was defined exploiting previous knowledge about the problem and physics-derived information.
4. Performances Evaluation: Comparison of the achieved performances by FFNN and 1D-CNN over the five regression tasks, by means of predictions vs. true labels and residuals plots.

Figure 1 shows the schematic overview of these four main steps. The detailed description of each step is described in the sections 2.1, 2.2, 2.3 and 2.4.

2.1 Dataset Preparation

In this work, a total of 22 olive oil samples were analysed. They were obtained from the 2019-2020 harvest and provided by the producer Conde de Benalúa, Granada, Spain. Oils quality assessment was performed by the producer according to the current European regulation for the commercial classification into EVOO, virgin olive oil (VOO), and lampante olive oil (LOO) categories.[3] The fluorescence spectrum of each olive oil was acquired using a miniaturized low-cost sensor, which was described in detail in a previous work.[11] Visual examples of raw fluorescence spectra, a complete description of samples preparation, and the details of the dataset can be found in the previously cited study. The 22 samples were measured 20 times this resulting in a total of 440 spectra. The final dataset, therefore, is made of 440 arrays, each having 1024 values, each corresponding to a pixel of the detector inside the sensor. The signal at each pixel represents the measured fluoresce intensity after the background subtraction and normalisation to have an average of zero and a standard deviation of one were applied. This normalisation was applied to make the input data small enough to avoid numerical problems during the training phase of the neural networks.[12] Moreover, five different continuous labels (chemical properties of olive oils) were associated with each oil array: acidity (%), peroxide value ($mEqO_2/kg$), K_{270}, K_{232} and ethyl esters (mg/Kg). The complete list of chemical properties values for each oil can be found in a previous work.[13]

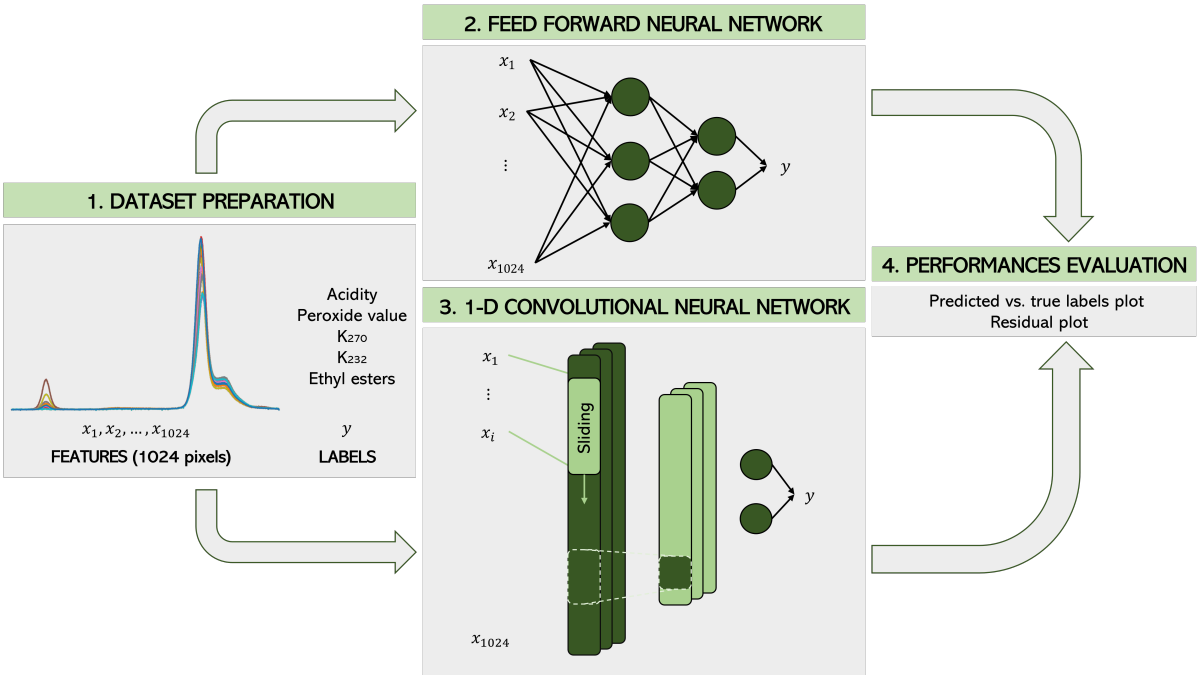

Figure 1: The four basic steps performed in this study: dataset preparation, feed-forward neural networks, 1-D convolutional neural networks, and comparison among the performances of the two types of architecture.

2.2 Feed Forward Neural Networks

Five different FFNN models[12] were trained to predict the five continuous labels associated to each oil (as described in Section 2.1), thus implementing a regression task.

The best combination of hyperparameters for each FFNN (namely, the number of layers, the number of neurons per layer, the number of epochs and the mini-batch size) was selected through a grid-search approach.[12] The evaluated values for each of the above-mentioned hyperparameters were:

- number of layers: 1, 3, 5, 6;
- number of neurons per layer: 8, 16, 32, 64;
- number of epochs: 1000, 2000, 3000, 5000, 10000, 20000;
- mini-batch size: 64, 128, 256.

The original dataset was split into training, validation, and test datasets, with relative proportions 60:20:20. The final set of hyperparameters was the one which maximised the mean absolute error (MAE) over the validation dataset, without overfitting the test dataset. MAE (as reported in Equation 1) was chosen as optimizing metric since it provided an average error score for each regression task:

$$MAE(X) = \frac{1}{|X|} \sum_{X} |y_{\text{predicted}} - y_{\text{measured}}| \tag{1}$$

where $|X|$ is the cardinality (i.e., the number of elements) of the dataset X, and y is the considered chemical property to be predicted. The optimization process was repeated separately for each chemical parameter. For all hidden neurons, the rectified linear unit (ReLU) function[12] was used as activation function (Equation 2):

$$\text{ReLU(x)} \equiv \max\{0, x\} \tag{2}$$

while the identity function was applied to the last layer, to perform the regression task.[12] The loss function used was the mean square error (MSE),[12] as reported in Equation 3:

$$MSE(X) = \frac{1}{|X|} \sum_{X} (y_{\text{predicted}} - y_{\text{measured}})^2 \tag{3}$$

where again $|X|$ is the cardinality (i.e., the number of elements) of the dataset X and y is the considered chemical property to be predicted. The FFNN was trained using the optimizer Adaptive Moment Estimation (Adam)[14] and the code was implemented using the TensorFlow ^{TM} Python library. All the models were trained with backpropagation.[15]

2.3 1-D Convolutional Neural Networks

The 1D-CNN[16–18] has a simple architecture which exploits physics-informed knowledge on the regression task. More specifically, since the spectrometer has a resolution of the order of 30 pixels, 40 pixels filter size in the first convolutional layer was an informed choice, due to the fact that changes in the signal in intervals shorter than 30 pixels may be simply due to noise. The same network was used for all the five chemical parameters. The network is made of one convolutional layer (6 filters of size 40), one max pooling layer (size 8), one convolutional layer (4 filters of size 20), one dropout layer (rate 0.5), two subsequent dense layers (4 neurons) and a final neuron for regression. As for the FFNN case, for all hidden neurons, the ReLU function[12] was used as activation function (Equation 2), while the same ReLU function was applied to the last layer, to perform the regression task. The loss function used was the MSE, as reported in Equation 3. The 1D-CNN was trained using the optimiser Adam[14] and the code was implemented using the TensorFlow ^{TM} Python library. All the models were trained with backpropagation.[15] For more details on the architecture of the 1D-CNN and hyperparameter tuning procedure the reader is referred to the work.[13]

2.4 Performances Evaluation

The chosen best model for each chemical parameter after FFNN hyperparameter tuning was compared to the 1D-CNN model (with the same number of epochs and mini-batch size as in the FFNN correspondent case). To evaluate the performance of an ANN the dataset is usually split into two parts, for example 80%, used as training dataset, and 20% used for validation.[12] Given the limited amount of data available for this study a leave-one-out cross-validation (LOOCV) was performed[19] for both the FFNN and 1D-CNN. Therefore, for each run one single oil sample was left out from the training and used for the validation as dataset. The final estimated MAE is then calculated averaging on all the oils.

3. RESULTS AND DISCUSSION

As described in Section 2.2, first the FFNN with best berformance was identified. For this reason the hyperparameter tuning was performed for each of the five chemical parameters to predict.

	Epochs	Layers	Neurons	Mini-batch size
Acidity	1000	1	8	64
Peroxide value	10000	5	16	128
K_{270}	10000	1	16	64
K_{232}	1000	1	8	64
Ethyl esters	1000	3	8	256

Table 1: Hyperparameters for the FFNN for the prediction of acidity, peroxide value, K_{270}, K_{232} and ethyl esters.

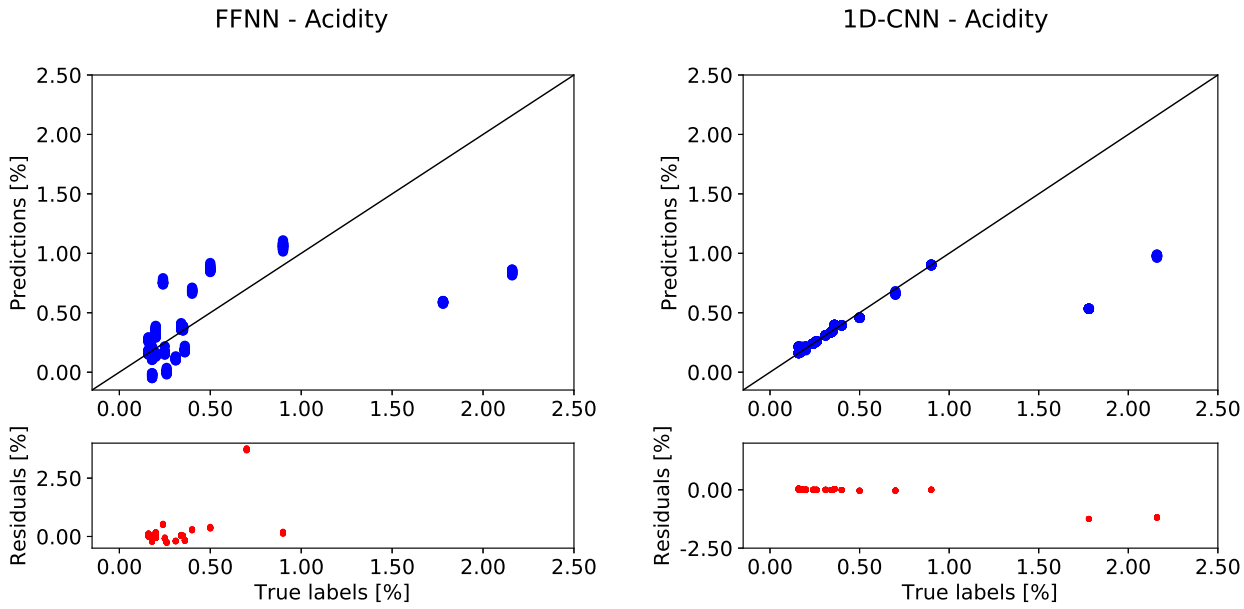

Figure 2: Comparison of the prediction performances for acidity using FFNN (left) and 1D-CNN (right). In each Figure, on the top, predictions against true labels are plotted, while on the bottom, the corresponding Residual for each predicted point are depicted. Each point inside the plots corresponds to an oil sample. The solid line corresponds to predictions equal to the true labels.

Table 1 reports the final best identified FFNN hyperparameters (number of epochs for model's training, number of network's layers, number of neurons for each single layer and mini-batch size used during the training phase) for the prediction of acidity, peroxide value, K_{270}, K_{232} and ethyl esters. Epochs represent the number of epochs used during the network's training phase, Layers the number of layers of the FFNN, Neurons the number of neurons in each layer and Mini-batch size the number of samples in each mini-batch used during the training phase. All the layers have the same number of neurons.

To visualise and better compare the performance of the two architecture, the predicted values are plotted against the measured (true labels) ones for each parameter. Additionally, the difference between the Predicted and the Actual value of the chemical property for each oil is plotted as Residuals below each curve.

Figure 2 shows the performances of the optimized FFNN model (left) and of the 1D-CNN model (right) for the acidity prediction, using the LOOCV approach described in 2.4. In each case, the predicted values (i.e. the model's outputs) are plotted against the true labels. Each point is an oil sample. The more points stand on the identity function, the better the corresponding prediction is. Moreover, for each point, the corresponding residual (i.e. the difference between the predicted and the true values) is also shown immediately below the predictions plot.

The results obtained for the prediction of the peroxide value of the olive oils are reported in Figure 3. Those for the UV-spectroscopy parameters K_{270} and K_{232} results in Figure 4, ethyl esters in Figure 5.

As it can be seen from Figures 2 to 5 the FFNN struggles to predict the chemical parameters correctly. Particularly for K_{270} and for the ethyl esters the FFNN does not seem to learn. The 1D-CNN, on the other hand, successfully predicts the chemical parameters and clearly outperforms the FFNN. In a few cases (Figure 2 and Figure 3) there are two and three samples respectively whose values are not predicted correctly. This should, however, not surprise because it is due to scarcity of the dataset. For the acidity, for example, there are only two labels above 1% and therefore not enough for the ANN to learn above this value.

To summarise the performance, the MAE for the prediction of each parameter for both types of ANN are reported in Table 2. This table enforces the findings that 1D-CNN always outperforms FFNN, in the prediction

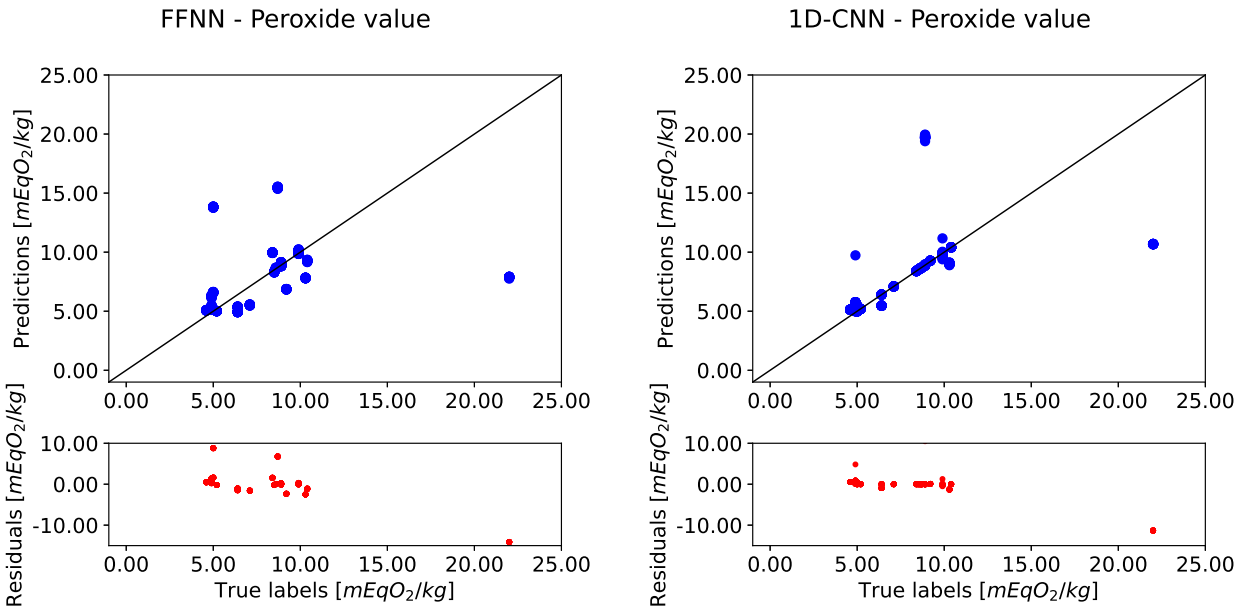

Figure 3: Comparison of the prediction performances for peroxide value using FFNN (left) and 1D-CNN (right). In each Figure, on the top, predictions against true labels are plotted, while on the bottom, the corresponding Residual for each predicted point are depicted. Each point inside the plots corresponds to an oil sample. The solid line corresponds to predictions equal to the true labels.

	MAE_{FFNN}	MAE_{1D-CNN}	σ_{FFNN}	σ_{1D-CNN}
Acidity	0.38	0.12	0.76	0.35
Peroxide value	1.94	1.31	3.28	3.19
K_{270}	0.016	0.010	0.014	0.013
K_{232}	0.054	0.039	0.061	0.039
Ethyl esters	7.1	3.6	7.9	4.3

Table 2: Overview of the performance of the two types of ANN for all the chemical parameters.

of each of the five olive oil chemical parameters. For each parameter the MAE and its standard deviation are always lower with 1D-CNN than with the correspondent FFNN.

4. CONCLUSION

In this work, five olive oils chemical parameters (acidity, peroxide value, K_{270}, K_{232} and ethyl esters) were predicted by applying two types of ANN, namely FFNN and 1D-CNN, to fluorescence spectra to fully characterise olive oil according to the European regulations. The main advantages of this approach are:

- the possibility of using non-destructive optical measurements which can be performed on low-cost and miniaturised optical instruments (like the one presented in[11]) instead of time-consuming and costly chemistry-based processes, like chromatography;
- the easier handling of samples, that do not require any preparation or dilution;
- the lack of pre-preprocessing of fluorescence spectra, that can be directly used as input to the neural network.

The results show that the 1D-CNN outperforms the FFNN, even if the FFNN architecture was optimised, while the 1D-CNN one was chosen as simple as possible, employing previous physics-based knowledge on the

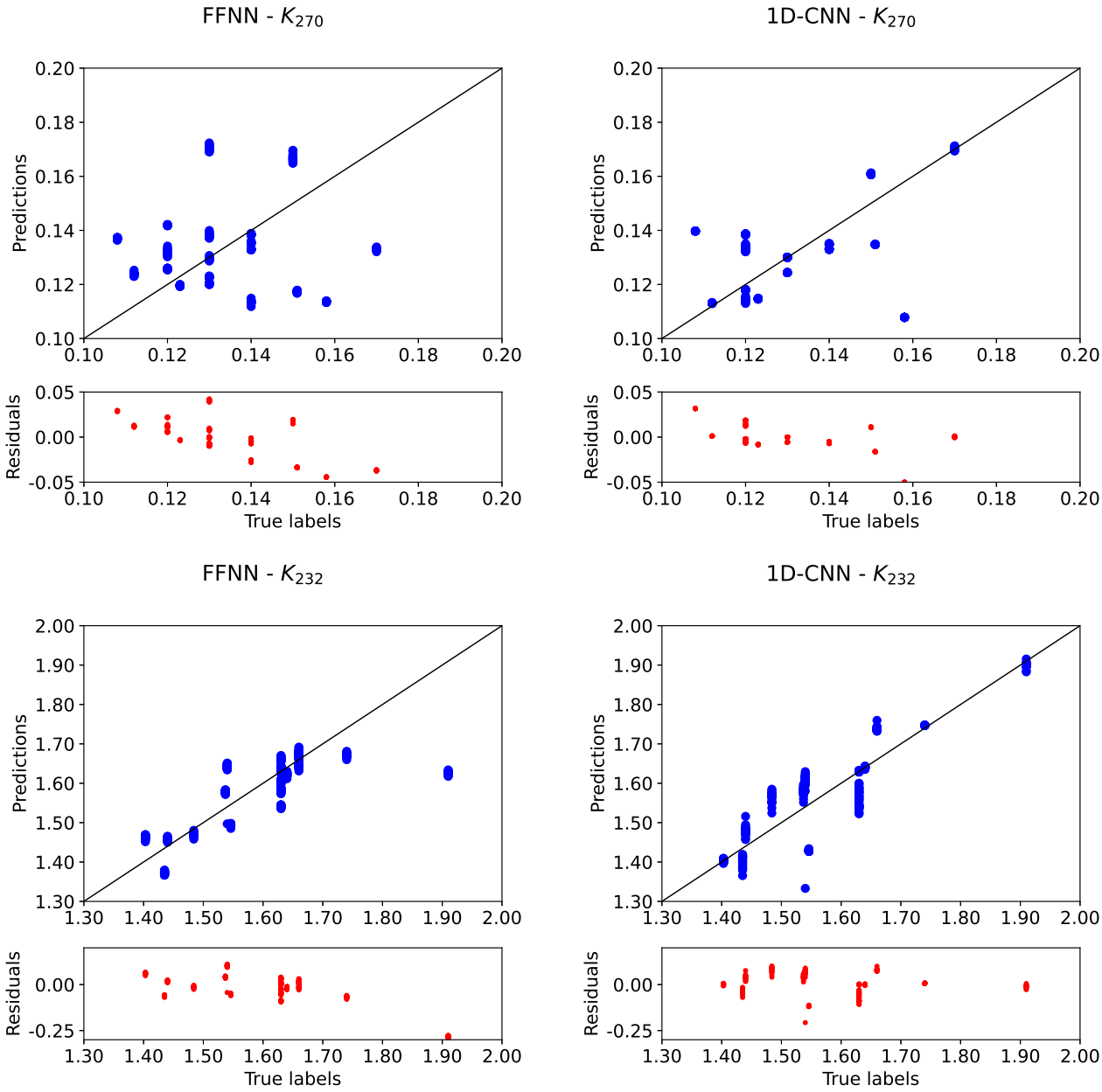

Figure 4: Comparison of the prediction performances for UV-spectroscopy parameters K_{270} and K_{232} using FFNN (left) and 1D-CNN (right). In each Figure, on the top, predictions against true labels are plotted, while on the bottom, the corresponding Residual for each predicted point are depicted. Each point inside the plots corresponds to an oil sample. The solid line corresponds to predictions equal to the true labels.

task. This behaviour may be due to the fact that the 1D-CNN, by means of filters, averages the information coming from several nearby pixels, thus considering not only a single pixel which is then weighted by a coefficient as in the FFNN, but the nearest neighbourhood of a single point. This is crucial when working with signals, where every feature is directly related not only in space, but also in time with the others. This finding enforces the use of ANN for parameters prediction from fluorescence spectra and its use in the olive oil sector, electing 1D-CNN as the best neural network approach to use with signals of this type.

The main limitation of this study is the reduced size of the dataset, which was however statistically addressed

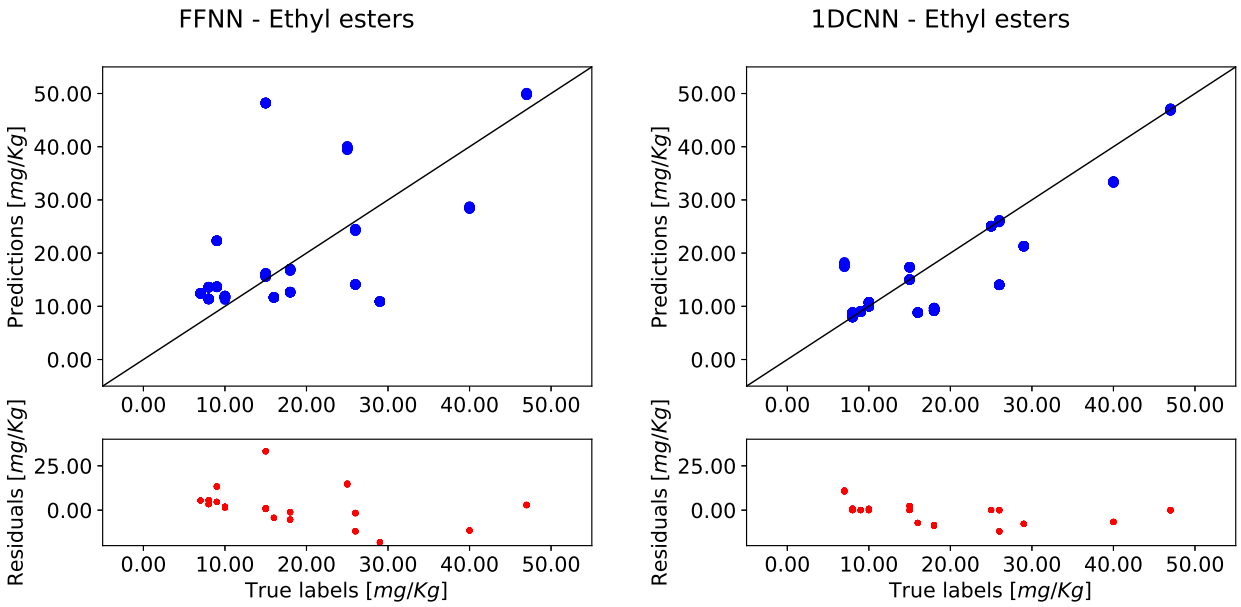

Figure 5: Comparison of the prediction performances for ethyl esters using FFNN (left) and 1D-CNN (right). In each Figure, on the top, predictions against true labels are plotted, while on the bottom, the corresponding Residual for each predicted point are depicted. Each point inside the plots corresponds to an oil sample. The solid line corresponds to predictions equal top the true labels.

applying LOOCV approach to evaluate the performances. Future development will consist in testing the proposed approach on larger olive oil datasets.

ACKNOWLEDGMENTS

The research was supported by the projects: "VIRTUOUS" funded by the European Union's Horizon 2020 Project H2020-MSCA-RISE-2019 Grant No. 872181, and "SUSTAINABLE" funded by the European Union's Horizon 2020 Project H2020-MSCA-RISE-2020 Grant No. 101007702.

REFERENCES

[1] Kiritsakis, A. and Markakis, P., "Olive oil: a review," *Advances in food Research* **31**, 453–482 (1988).

[2] Georgiou, C. A. and Danezis, G. P., [*Food authentication: Management, analysis and regulation*], John Wiley & Sons (2017).

[3] "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**, 1–83 (1991).

[4] Guzmán, E., Baeten, V., Pierna, J. A. F., and García-Mesa, J. A., "Evaluation of the overall quality of olive oil using fluorescence spectroscopy," *Food Chemistry* **173**, 927–934 (2015).

[5] Merás, I. D., Manzano, J. D., Rodríguez, D. A., and de la 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* **178**, 751–762 (2018).

[6] Kongbonga, Y. G. M., Ghalila, H., Onana, M. B., Majdi, Y., Lakhdar, Z. B., Mezlini, H., and Sevestre-Ghalila, S., "Characterization of vegetable oils by fluorescence spectroscopy," *Food and Nutrition Sciences* **2**(7), 692–699 (2011).

[7] Sikorska, E., Khmelinskii, I., and Sikorski, M., "Analysis of olive oils by fluorescence spectroscopy: methods and applications," *Olive oil-constituents, quality, health properties and bioconversions* , 63–88 (2012).

[8] Guimet, F., Boqué, R., and Ferré, J., "Cluster analysis applied to the exploratory analysis of commercial spanish olive oils by means of excitation- emission fluorescence spectroscopy," *Journal of agricultural and food chemistry* **52**(22), 6673–6679 (2004).

[9] Poulli, K. I., Mousdis, G. A., and Georgiou, C. A., "Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy," *Analytica Chimica Acta* **542**(2), 151–156 (2005).

[10] Gonzalez-Fernandez, I., Iglesias-Otero, M., Esteki, M., Moldes, O., Mejuto, J., and Simal-Gandara, J., "A critical review on the use of artificial neural networks in olive oil production, characterization and authentication," *Critical reviews in food science and nutrition* **59**(12), 1913–1926 (2019).

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

[12] Michelucci, U., [*Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks*], APRESS Media, LLC (2018).

[13] Venturini, F., Sperti, M., Michelucci, U., Gucciardi, A., Martos, V. M., and Deriu, M. A., "Physico-chemical properties extraction from the fluorescence spectrum with 1d-convolutional neural networks: application to olive oil," *preprint (arXiv)* (2022).

[14] Kingma, D.P.; Ba, J. A., "Adam: A method for stochastic optimization. in proceedings of 3rd.," *In Proceedings of 3rd International Conference on Learning Representations, ICLR 2015* , 1–15 (2015).

[15] Kelley, H. J., "Gradient theory of optimal flight paths," *Ars Journal* **30**(10), 947–954 (1960).

[16] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L., "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems* **2** (1989).

[17] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* **86**(11), 2278–2324 (1998).

[18] Michelucci, U., [*Advanced applied deep learning: convolutional neural networks and object detection*], APRESS Media, LLC (2019).

[19] Michelucci, U. and Venturini, F., "Estimating neural network's performance with bootstrap: A tutorial," *Machine Learning and Knowledge Extraction* **3**(2), 357–373 (2021).