Understanding the learning mechanism of convolutional neural network applied to fluorescence spectra

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# Understanding the learning mechanism of convolutional neural networks applied to fluorescence spectra

Francesca Venturini, Umberto Michelucci, Michela Sperti, Arnaud Gucciardi, Marco Deriu

Francesca Venturini, Umberto Michelucci, Michela Sperti, Arnaud Gucciardi, Marco A. Deriu, "Understanding the learning mechanism of convolutional neural networks applied to fluorescence spectra," Proc. SPIE 12438, AI and Optical Data Sciences IV, 124380O (15 March 2023); doi: 10.1117/12.2647809

SPIE.

Event: SPIE OPTO, 2023, San Francisco, California, United States

# Understanding the learning mechanism of convolutional neural network applied to fluorescence spectra

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

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

## ABSTRACT

The power of artificial neural networks to determine the quality and properties of olive oil was proven by several studies in the last years. Less clear is, however, how the neural network is able to extract useful information from the input data. This work investigates the learning mechanism of one-dimensional convolutional neural networks (1D-CNNs) trained to predict the physicochemical properties of olive oil from single fluorescence spectra. Such a 1D-CNN can successfully predict the parameters relevant to the quality assessment: acidity, peroxide value, and UV absorbance. To go beyond a simple quality assessment algorithm, it is important to identify which spectral features in the measured spectra are correlated with each chemical parameter and therefore with the quality of olive oil. To obtain this information, explainability techniques can be used by studying the latent feature space generated by the intermediate layers of the one-dimensional trained convolutional neural network. This work analyses in detail the common features that are used by the 1D-CNN to predict the two physicochemical parameters: acidity and $K_{232}$.

**Keywords:** Fluorescence spectroscopy; olive oil; machine learning; artificial neural networks; quality control; explainability; convolutional neural networks

## 1. INTRODUCTION

Artificial neural networks have shown significant capabilities in a wide range of fields, including fluorescence and Raman spectroscopy[1–3] and are known for the undiscussed capability to automatically extract relevant features from data,[4,5] eliminating the need for manual preprocessing. However, the "black-box" nature of these models[6] is a point of criticism that undermines the credibility of the results and limits the full exploitation of their potential for practical applications. Even when neural networks are able to predict certain quantities with high accuracy, it is often unclear which input information is used to make the prediction, and whether it is possible to extract this information from the model itself. Additionally, understanding which features are relevant in the input data can provide new insights into the underlying physical or chemical processes.

In spectroscopic applications of neural networks the input data are one or two dimensional long arrays of real numbers, representing, for example, the light intensity at different wavelengths. Depending on the spectroscopic characteristics, the relevant features may be more or less broad, therefore, covering certain ranges and not individual pixels. This makes it difficult to apply standard feature importance analysis methods, such as forward or backward selection.[7] Various approaches have been proposed to overcome this challenge, such as interval

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

AI and Optical Data Sciences IV, edited by Bahram Jalali, Ken-ichi Kitayama,
Proc. of SPIE Vol. 12438, 124380O · © 2023 SPIE
0277-786X · doi: 10.1117/12.2647809

partial least squares,[8] genetic algorithms,[9] among others, but all have been criticized for various downsides[10,11] such as low reproducibility, overfitting and the loss of important features.

This work presents an in-depth study of the latent features of one-dimensional convolutional neural networks (1D-CNNs) trained to extract two chemical parameters from raw spectra of olive oil samples. This work expands on a previous preliminary work in which 1D-CNN networks were used to extract the UV-spectroscopy parameter $K_{232}$ adding an additional one, the acidity[12] and analyzing and comparing results on the entire dataset composed of multiple olive oil samples. The output of the first convolutional layer (the feature maps) is analyzed and visualized to show which spectroscopic features (identified by wavelength ranges) are considered relevant by the network for the final prediction. To achieve this, feature maps are compared with the raw spectra, which highlight wavelength windows that the trained neural network considers important. Since these spectral regions are common to all the oils, it can be concluded that the learning is robust and that the 1D-CNN identifies physically significant regions of the fluorescence spectrum successfully.

The main contributions of this work are two: the study of latent features for a complete dataset of olive oil samples, and the identification, discussion and comparison of the wavelength ranges identified as relevant by the trained models for the two parameters $K_{232}$ and acidity.

## 2. MATERIAL AND METHODS

### 2.1 Dataset

The dataset used for this work contains fluorescence spectra and chemical parameters of 24 olive oils from the 2019-2020 harvest and provided by the Conde de Benala producer, Granada, Spain. The fluorescence spectrum of each olive oil was acquired using a miniaturized low-cost sensor, which was described in detail in a previous work.[13] From the raw fluorescence spectra only the background (dark counts) was subtracted, and then the spectra were normalized. Therefore, the signal at each pixel is the measured fluoresce intensity after the background subtraction, and normalisation to have an average of zero and a standard deviation of one was applied. This normalisation was applied to make the values of the input data small enough to avoid numerical problems during the training phase of neural networks.[14] All the fluorescence measurements in this work were made on undiluted samples under identical conditions.

The quality evaluation according to the current European regulation for the commercial classification into EVOO, virgin olive oil (VOO), and lampante olive oil (LOO) categories[15] was performed by the producer. The chemical parameters were determined by accredited laboratories using the procedures described in the European Commission regulation and its amendment.[15,16]

For the list of oil samples, their chemical and quality characteristics the reader is referred to the work.[17] The dataset is publicly available available.[18]

### 2.2 Convolutional Neural Network

The 1D-CNN architecture was previously described[13] and will only be minimally discussed here for clarity. The 1D-CNN consists 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 output neuron for regression. The network is shown in Figure 1. Apart from the output neuron that uses the identity activation functions, all other layers use the Rectified Linear Unit (ReLU) activation function. The network architecture can be seen in 1 and details in the previous paper.[13] The size of the filters was chosen to be 40 pixels, slightly larger than the spectrometer resolution.

The loss function used for the training is the mean squared error (MSE),[14] as reported in Equation 1, where $y^{[i]}_{\text{predicted}}$ is the predicted and $y^{[i]}_{\text{measured}}$ is the measured parameter value of the $i^{th}$ input respectively.

$$\text{MSE}(X) = \frac{1}{N}\sum_{i=0}^{N}(y^{[i]}_{\text{predicted}} - y^{[i]}_{\text{measured}})^2. \quad (1)$$

$N$ indicates the size of the dataset on which the metric is being evaluated. The 1D-CNN was trained using the optimiser Adam[19] and the code was implemented using the TensorFlow $^{TM}$ Python library. All the models were

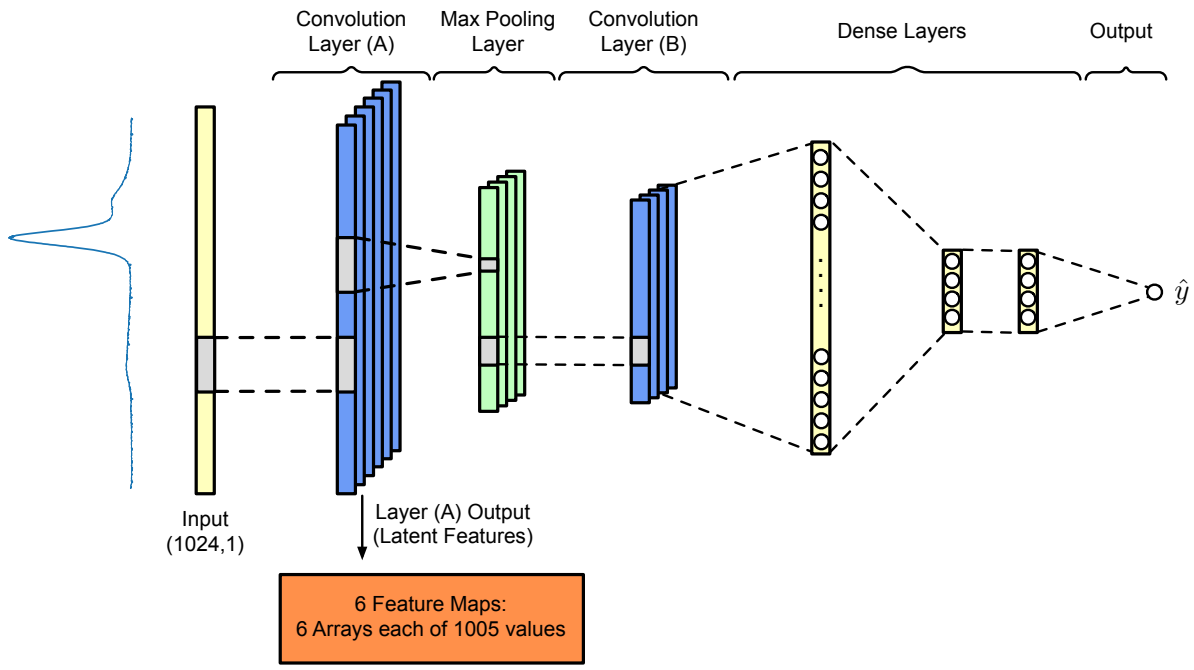


Figure 1. Schematic representation of the 1D-CNN used in this paper. The blue layers are convolutional ones, the green max pooling layers and the yellow marked ones are dense layers. The output layer has 1 neuron with the identity activation function. The orange box indicates the output of the first convolutional layer (indicated with (A)), also called latent features. The analysis of the those arrays and their visualisation is used in this work to study the learning of the network. The output of layer (A) consist of 6 arrays, each of 1005 values.

trained with back-propagation. The network was trained for 5000 epochs with a mini-batch size of 64. Due to the small size of the data set, a leave-one-out cross-validation approach was used.[20] This means that at the end 22 models will be trained, each trained on a slightly different training dataset.

### 2.3 Explainability Approach

The explainability approach for the neural network involves understanding and interpreting the inner workings of the model, particularly with regards to how it processes input data and produces output predictions. One way to achieve this is through visualizing latent features, which are the underlying representations learned by the model during training. In particular, this work considers the output of the first convolutional layer, that is indicated with (A) in Figure 1. The output consist of 6 arrays, each having 1005 values.

The visualisation was realized with the following steps.

1. All the outputs of the (A) convolutional layer were concatenated. This results in a 2-dimensional array of (1005) arrays.
2. The arrays were sorted in ascending order of the position of the maximum of each array.
3. The sorted array were then used to generate a heatmap.

The results for the two parameters analyzed, the acidity and the $K_{232}$, two figures are shown. One obtained with the unmodified values, and the other re-scaled with the $\tanh(\cdot)$ function. The $\tanh(\cdot)$ is equal to one for large values of its input, and thus normalizes all the arrays to have values less than one. As discussed in Section 3, the use of the scaling with the $\tanh(\cdot)$ makes it easier to identify regions where the latent features are significantly different than zero.

## 3. RESULTS AND DISCUSSION

The raw fluorescence spectra of selected EVOOs, VOOs, and LOOs under excitation at 365 nm and 395 nm are shown in Figure 2. The spectra in all panels are just one single spectrum with the dark background subtracted, without averaging or smoothing. In the figure, the dashed lines mark the maxima of the strong chlorophyll contributions.

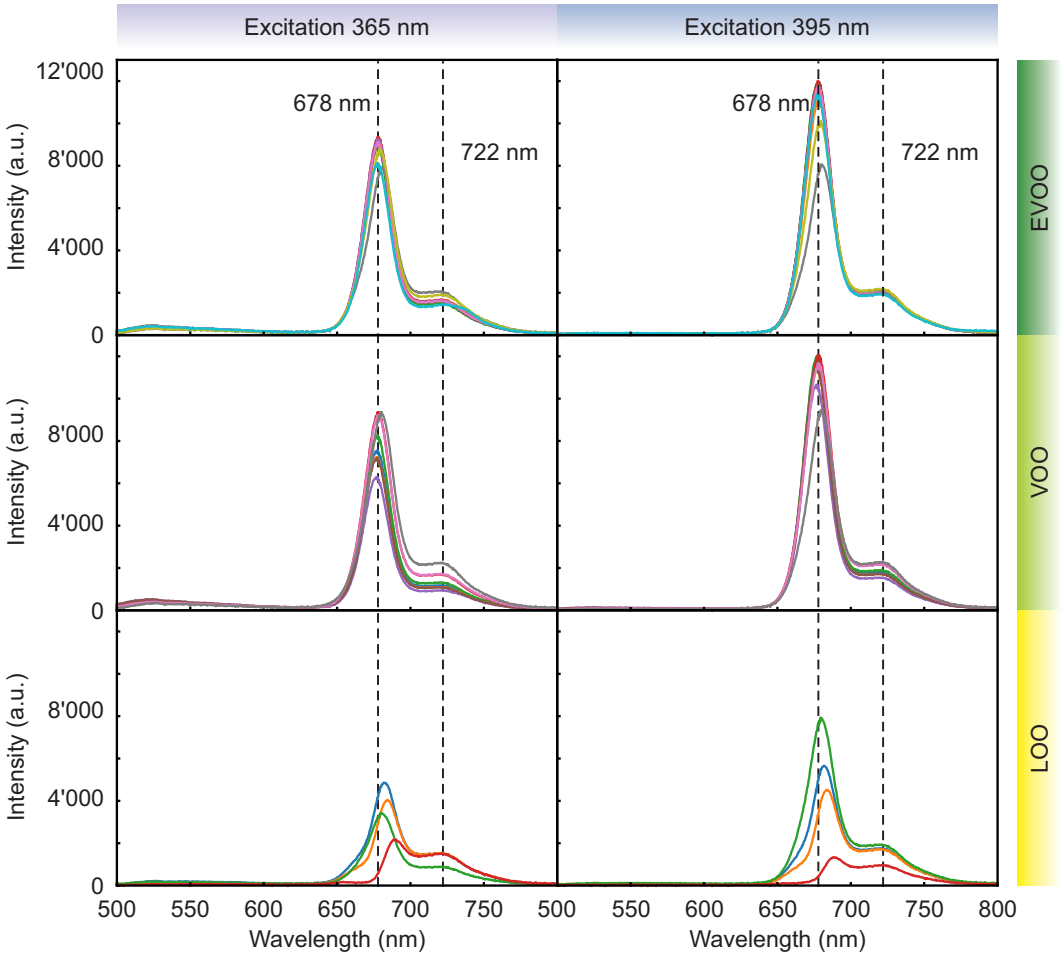


Figure 2. Fluorescence emission spectra of the measured olive oils excited at 365 nm and 395 nm nm divided in the quality classes EVOO, VOO and LOO. Each curve shows a single spectrum without averaging or smoothing. Reproduced from.[17]

The heatmap of the latent features realized as described in Section 2.3 is shown in Figure 3. The heatmaps are obtained with the latent features from the models trained to predict the acidity and $K_{232}$ parameter (in the top and bottom panels respectively). The left panels are obtained with unmodified data, and the right ones with the data transformed with tanh(·). In Figure 3 the wavelengths of the excitation LED and of the two spectral maxima at 678 nm and 722 nm are also marked to facilitate the interpretation of the results.

The analysis shows that there are mainly four relevant spectral regions that are schematically illustrated in Figure 4.

- The region from up to ca. 640 nm and above 750 nm (indicated with (A) in Figure 4): in this region the latent features are not so intense (as it can be seen in left panel of Figure 3) indicating that the networks do not consider this spectral region very useful for the predictions.
- The region immediately to the left of the first peak (between ca. 640 nm and 678 nm) (indicated with (B) in Figure 4): in this region the latent features have the highest intensity, indicating that the networks

consider this spectral window very important. This region can be interpreted as the ascending part of the first peak.

- The region between the first peak at 678 nm and the second one at 722 nm (indicated with (C) in Figure 4): a portion of the latent features have also a high intensity. This region can be interpreted as the descending part of the first peak and the shoulder up to the second peak (see Figure 2).
- The region between the second peak at 722 nm and ca. 750 nm (indicated with (D) in Figure 4): this part coincides with the descending part of the second peak. Also this region is considered important.

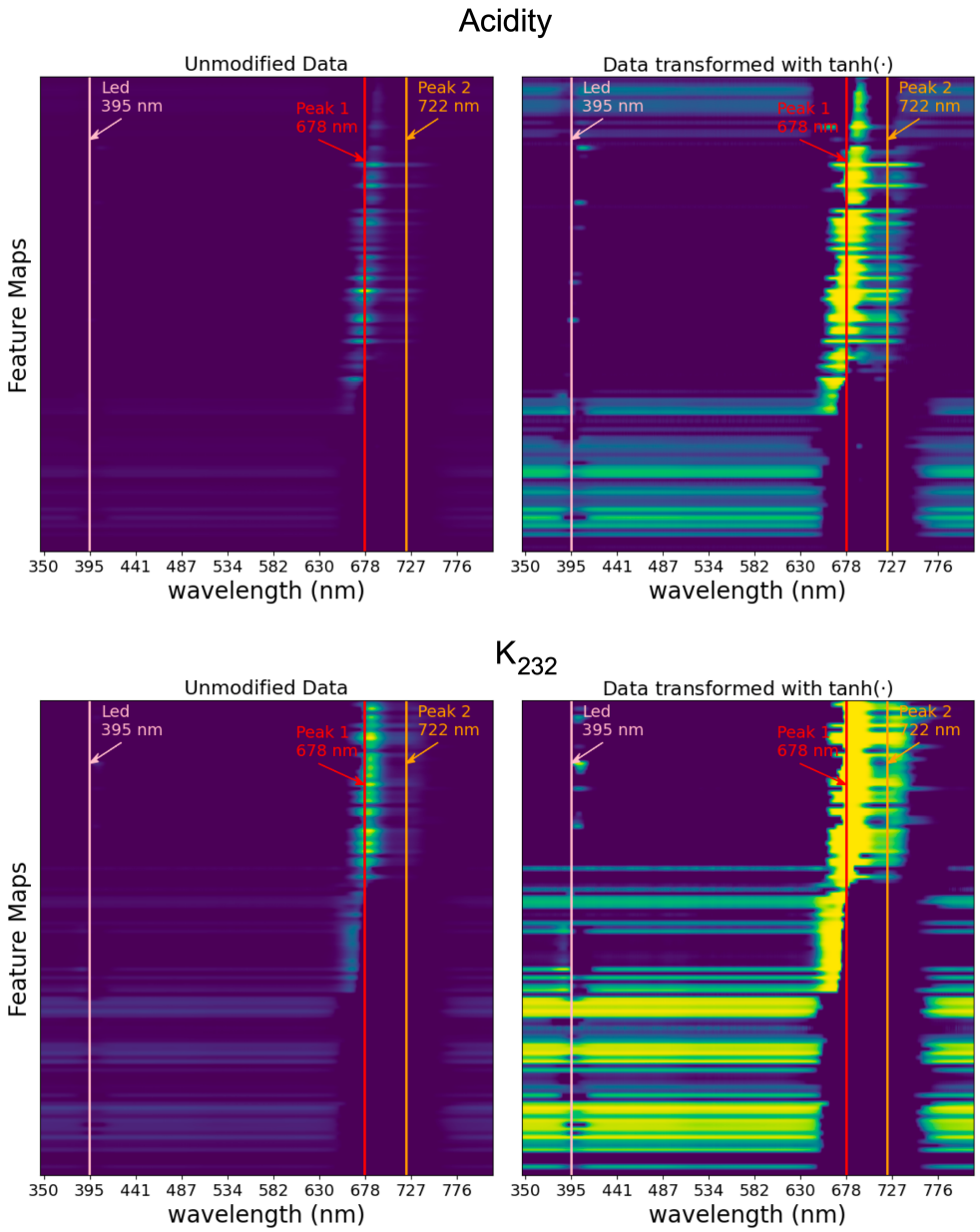


Figure 3. Latent features heatmap for the models trained to predict the acidity parameter. The left panel shows the heatmap obtained with the unmodified data. The one on the right the heatmap obtained by transforming the data with the function tanh(·). Note that the $x$-axis is not linear in nm since the raw data are arrays of pixels that are then converted in wavelengths with a cubic function.

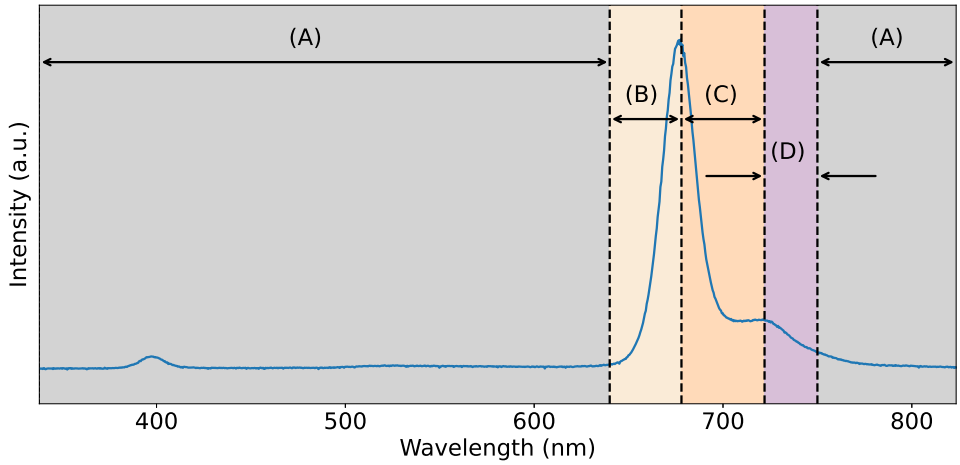


Figure 4. An overview of relevant spectral regions used by the networks for the prediction of the two physicochemical parameters acidity and $K_{232}$ and a typical fluorescence spectrum of olive oil.

It is interesting to see how neural networks that are trained to predict the $K_{232}$ parameter tend to give much less importance to the spectral region after the first peak (after 678 nm). To predict the acidity, at least according to the trained networks, the spectral region with wavelengths greater than 678 nm is more important than the same region to predict the $K_{232}$ parameter. This indicates that the trained networks use different spectral regions in different ways, also indicating that different spectral windows contains different information correlated to different chemical parameters.

## 4. CONCLUSION

In conclusion, studying the latent features of the trained 1D-CNN, allows to identify the relevant spectral features with explainability and visualisation techniques. Even if the dataset is small using the leave-one-our validation method it is possible to obtain robust results. Not only that, but the approach described in this work allows to check if all the different trained models (remember that in leave-one-out one trains as many models as different observations) identify the same spectral regions, increasing the trust in the relevance of the identified spectral regions. Networks trained to predict different chemical parameters identify the same main spectral regions, but give different weights to them. The fact that models trained to perform different tasks identify the same spectral window (although with different weights) gives us a strong evidence that the regions are the ones that contains the relevant information correlated to the different chemical properties of the olive oil.

## ACKNOWLEDGMENTS

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.

## REFERENCES

[1] Yuanyuan, C. and Zhibin, W., "Quantitative analysis modeling of infrared spectroscopy based on ensemble convolutional neural networks," *Chemometrics and Intelligent Laboratory Systems* **181**, 1–10 (2018).

[2] Acquarelli, J., van Laarhoven, T., Gerretzen, J., Tran, T. N., Buydens, L. M., and Marchiori, E., "Convolutional neural networks for vibrational spectroscopic data analysis," *Analytica chimica acta* **954**, 22–31 (2017).

[3] Zhang, X., Lin, T., Xu, J., Luo, X., and Ying, Y., "Deepspectra: An end-to-end deep learning approach for quantitative spectral analysis," *Analytica chimica acta* **1058**, 48–57 (2019).

[4] Malek, S., Melgani, F., and Bazi, Y., "One-dimensional convolutional neural networks for spectroscopic signal regression," *Journal of Chemometrics* **32**(5), e2977 (2018).
[5] Ni, C., Wang, D., and Tao, Y., "Variable weighted convolutional neural network for the nitrogen content quantization of masson pine seedling leaves with near-infrared spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **209**, 32–39 (2019).
[6] Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B., "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences* **116**(44), 22071–22080 (2019).
[7] James, G., Witten, D., Hastie, T., and Tibshirani, R., [*An introduction to statistical learning*], vol. 112, Springer (2013).
[8] Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., and Engelsen, S. B., "Interval partial least-squares regression (i pls): A comparative chemometric study with an example from near-infrared spectroscopy," *Applied spectroscopy* **54**(3), 413–419 (2000).
[9] Arakawa, M., Yamashita, Y., and Funatsu, K., "Genetic algorithm-based wavelength selection method for spectral calibration," *Journal of Chemometrics* **25**(1), 10–19 (2011).
[10] Yun, Y.-H., Bin, J., Liu, D.-L., Xu, L., Yan, T.-L., Cao, D.-S., and Xu, Q.-S., "A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration," *Analytica chimica acta* **1058**, 58–69 (2019).
[11] Zhang, J., Yan, H., Xiong, Y., Li, Q., and Min, S., "An ensemble variable selection method for vibrational spectroscopic data analysis," *RSC advances* **9**(12), 6708–6716 (2019).
[12] Venturini, F., Michelucci, U., Sperti, M., Gucciardi, A., and Deriu, M. A., "One-dimensional convolutional neural networks design for fluorescence spectroscopy with prior knowledge: explainability techniques applied to olive oil fluorescence spectra," in [*Optical Sensing and Detection VII*], **12139**, 326–333, SPIE (2022).
[13] 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).
[14] Michelucci, U., [*Applied Deep Learning - A Case-Based Approach to Understanding Deep Neural Networks*], APRESS Media, LLC (2018).
[15] "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).
[16] "Commission implementing regulation no 1348/2013 of december 17 2013," *Official Journal of the European Union* **338**, 31–67 (2013).
[17] Venturini, F., Sperti, M., Michelucci, U., Gucciardi, A., Martos, V. M., and Deriu, M. A., "Extraction of physicochemical properties from the fluorescence spectrum with 1d convolutional neural networks: Application to olive oil," *Journal of Food Engineering* **336**, 111198 (2023).
[18] Venturini, F., Sperti, M., Michelucci, U., Gucciardi, A., Martos, V., and Deriu, M. A., "Dataset of Fluorescence Spectra and Chemical Parameters of Olive Oils." https://data.mendeley.com/datasets/thkcz3h6n6/6 (Dec. 2022).
[19] 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).
[20] Michelucci, U. and Venturini, F., "Estimating neural network's performance with bootstrap: A tutorial," *Machine Learning and Knowledge Extraction* **3**(2), 357–373 (2021).