**Use of deep learning in combination with FT-NIR spectroscopy for the analysis of extra virgin olive oil**

Andrea Bandiera (andrea.bandiera@unitus.it)

Dept. DIBAF, University of Tuscia, Italy

Tutor: Prof. Roberto Moscetti

The first two years of the PhD project were dedicated to two main objectives i) acquiring chemical and spectral data from various categories of olive oils to develop predictive models for the chemical profile; and ii) developing NIR-based models to quantify adulterations of EVOO with seed oils. Data acquisition is currently ongoing, and preliminary results have shown the following: i) a well-distributed range of chemical profile features, which is essential for further model development, and ii) the successful calibration of Partial Least Squares (PLS) predictive models for adulteration, showing excellent performance metrics (R2 = 0.99; near-zero BIAS; and RMSE ranging from 1.7 to 3.2).

Fattibilità di impiego dell’apprendimento profondo in combinazione alla spettroscopia FT-NIR per l’analisi dell’olio extra vergine di oliva

Nei primi due anni di dottorato sono stati i) acquisiti dati spettrali e chimici da diverse categorie di olio di oliva per lo sviluppo di modelli predittivi del profilo chimico, e ii) sviluppati modelli basati sulla spettroscopia NIR per la quantificazione di adulteranti in EVOO. L’acquisizione dati è in corso. Dai risultati preliminari si evince i) una buona distribuzione dei parametri nel profilo chimico (necessaria per lo sviluppo di un buon modello) e ii) modelli predittivi basati su PLS con metriche prestazionali eccellenti (R2 > 0.99; BIAS tendente a 0; and RMSE tra 1.7 e 3.2).

**Keywords**: Olive oil, quality, adulteration, shelf-life, FT-NIR spectroscopy, chemometrics, deep learning

# 1. Introduction

In accordance with the PhD thesis project (Bandiera A., 2022), the present manuscript reports its main results.

1. **Development of the research strategy:** the study includes the state-of-the-art analysis for A1.1, which involves the use of UV-Vis and NIR spectroscopy for the quali-quantitative analysis of EVOOs. It also includes A1.2, the selection of classical analytical methods for the analysis of EVOOs, and A1.3, the selection and testing of DNN algorithms for the modelling of EVOOs’ spectral data.
2. **Chemical-profile and shelf-life prediction models:** this activity involves the development of DNN-based models to assess the chemical profile and shelf-life of EVOOs. It includes A2.1, shelf-life tests; A2.2, UV-Vis and FT-NIR spectral scans, and chemical analysis of samples; A2.3, calibration of models using both classical and deep chemometrics; and A2.4, validation of models using external datasets.
3. **Adulteration identification models:** this activity focuses on the identification and quantification of adulteration in EVOOs using DNN models. It includes A3.1, samples preparation and UV-Vis and FT-NIR scans; A3.2, calibration of predictive models; and A3.3, validation of predictive models.

# 2. Materials and Methods

**2.1 Sample preparation**

For activity A2, monthly olive oil samples from Umbria Olii International Spa, an olive oil refinery factory (Perugia, Italy) included extra virgin, virgin, lampante, pomace and refined. Chemical analytical reports were provided for each sample. Samples were stored in dark at 25 °C for 24 hours at 25 °C before spectral analysis. Samples for the activity A3 were prepared (Vanstone et al., 2018) by blending EVOO from the “Oleificio Sociale di Canino” (Viterbo, Italy) with 4 seed oils (i.e., peanut, sunflower, maize, soy) purchased from a local market. The blend concentrations for each adulterant were as follows: 0, 0.5, 1.5, 3, 5, 10, 15, 20, 60 and 100 % (w/w). Four biological replicates were prepared.

**2.2 Spectra acquisition**

Spectra were acquired using UV-Vis and FT-NIR spectrophotometers, mod. Lambda 850+ (PerkinElmer, USA) and mod. Antaris II (ThermoFisher, USA), respectively. The former operated within the spectral range of 380-900 nm with a resolution of 1 nm, while the latter covered the range of 1000-2500 nm with a resolution of 1.93 nm. Before scanning, the samples were filtered using fast paper qualitative filters. Spectral acquisition took place at a controlled temperature of approx. 32 °C using optical-glass cells with path lengths of 1 and 6 mm. Three technical replicates were carried out. Spectral acquisitions are still going, and currently, around 300 and 400 spectra were acquired for A2 and A3, respectively.

**2.3 Descriptive statistics**

The chemical-profile data from the analytical reports produced by Umbria Olii International (activity A2) were subjected to statistical analysis to verify deviations from the normal distribution for each predictand (i.e., analyte).

**2.4 Calibration and cross-validation of predictive models**

Partial Least Squares (PLS) models, developed using Matlab R2017b software (Mathworks, USA) and PLS toolbox v8.6.2 (Eigenvector research Inc., USA), were used for chemometrics modeling (Nallan Chakravartula S.S. et al. 2022). Deep chemometrics (DNN) models were developed using Python3 with Keras and PyTorch/TensorFlow packages. Spectral pre-treatments, including Standard Normal Variate (SNV); Multiplicative Scatter Correction (MSC); Extended MSC (EMSC); and Savitzky-Golay 1st and 2nd derivatives with a 2nd or 3rd order polynomial fitted over a window of 25-45 features, were tested alone or in combination. These pre-treatments were always followed by mean-centering (MC) or autoscaling (AS). The optimal number of latent variables (LVs) was selected using the “Venetian blind” cross-validation method. Performances were evaluated in terms of Root Mean Squared Error (RMSE), systematic error (BIAS), and R2.

# 3. Results and Discussion

Results of the analysis conducted for activities A2 and A3 are presented. It is important to note that the ongoing scans of UV-Vis and FT-NIR data limits the presentation to the statistical analysis of A2 data and the calibration of A3 models.

**3.1 Chemical-profile data overview**

Table 1 shows the descriptive statistics of a set of analytical data acquired by the refinery factory. Additional features (data not shown) include triolein, ester fraction (cholesterol, brassicasterol, campesterol, stigmasterol, delta7-stigmasterol, delta7-avenasterol), fatty acids (lauric, myristic, palmitic, palmitoleic, stearic, arachidic, eicosenoic, behenic, erucic and lignoceric acids), and trans-fatty acids (oleic, linoleic and linolenic acids). Data had a good range for each feature, but not all of them passed the normality tests, posing a potential risk to model performances. However, sampling is ongoing, and a more representative sample size should improve this aspect. In this context, data transformation will be also tested.

***Table 1*** *Descriptive statistics of some analytes used in the development of PLS and DNN chemical-profile prediction models.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Stats** | **Acidity** | **Peroxides** | **Ethyl esters** | **Waxes** | **Total sterols** | **Oleic ac.** | **Linoleic ac.** | **Linolenic ac.** |
| (% w/w) | (meq/kg) | (mg/kg) | (mg/kg) | (g/kg) | (%) | (%) | (%) |
| Min | 0.05 | 0.10 | 4.25 | 47.40 | 0.14 | 18.05 | 0.01 | 0.01 |
| Max | 15.20 | 12.30 | 27.94 | 542.40 | 11.20 | 76.89 | 0.50 | 0.10 |
| Mean | 3.17 | 7.98 | 12.95 | 238.20 | 1.68 | 68.62 | 0.06 | 0.04 |
| St. Dev. | 3.98 | 3.00 | 4.79 | 120.14 | 1.88 | 11.72 | 0.10 | 0.03 |

**3.2 Adulteration prediction models**

Table 2 shows the best predictive model for each adulterant selected among 140 spectral pre-treatments.

***Table 2*** *Summary of best PLS predictive models for adulterant quantification in EVOO.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Adulterant** | **Pre-treatment** | | | | **LVs**e | **RMSE (%)** | |  | **BIAS (%)** | |  | **R2** | |
| SGa | Deriv.b | SCc | Norm.d |  | Cf | CVg |  | C | CV |  | C | CV |
| peanut | 25 | 1st | EMSC | MC | 5 | 2.796 | 3.245 |  | -1.641E-12 | 0.108 |  | 0.989 | 0.985 |
| sunflower | 29 | none | EMSC | MC | 2 | 1.735 | 1.833 |  | 0E+00 | -0.039 |  | 0.996 | 0.995 |
| maize | 25 | 1st | EMSC | AS | 2 | 1.910 | 2.010 |  | -7.105E-15 | -0.014 |  | 0.995 | 0.994 |
| soy | 29 | none | EMSC | MC | 2 | 1.775 | 1.848 |  | 5.329E-14 | -0.004 |  | 0.996 | 0.995 |

*a, Savitzky-Golay filter; b, derivative; c, scatter correction; d, normalization; e, latent variables; f, calibration; f, cross validation.*

For all four models, the use of a SG smoothing window consisting of 25-29 features improved the signal-to-noise ratio, resulting in enhanced prediction performance. The models for peanut and maize oils exhibited notable improvement when subjected to the 1st derivative transformation. Additionally, the removal of light scattering through EMSC was crucial in all tests. Spectra normalisation using the MC proved effective in most cases, except for the maize model, which showed superior performance with AS normalization. Two LVs were captured over 95 % of the variance in all models, except for peanut model, which required five LVs. Notably, the peanut model showed a consistently higher RMSE, approx. 50 % greater than the other models. This difference can be attributed to a few possible outliers in the 60 % concentration adulteration. Further tests, including outlier analysis, are required to enhance the model precision. BIAS values across all models tended toward zero, indicating excellent accuracy. Future studies will involve the development of PLS models for the chemical profile and EVOO shelf-life tests, as well as the implementation of Deep Neural Networks in all case studies.

# 4. References

# Bandiera A. (2022). Use of deep learning in combination with FT-NIR spectroscopy for the analysis of extra virgin olive oil. *In proc.s of the 26th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology*, Asti (Italy), 19-21 September, 2022, pp. 33-34

# .Nallan Chakravartula S.S., Moscetti R., Bedini G., Nardella M., Massantini R. (2022) Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: *A case study on coffee, Food Control*. 135.

Vanstone, N., Moore, A., Martos, P., & Neethirajan, S. (2018). Detection of the adulteration of extra virgin olive oil by near-infrared spectroscopy and chemometric techniques. *Food Quality and Safety*, 2(4), 189–198.