The objective of this work was to evaluate the ability of artificial neural networks (ANN) in near infrared (NIR) spectra calibration models to predict the total polyphenolic content, antioxidant activity, and extraction yield of the olive leaves aqueous extracts prepared with three extraction procedures (conventional extraction, microwave-assisted extraction, and microwave-ultrasound-assisted extraction). Partial least squares (PLS) models were developed from principal component analyses (PCA) scores of NIR spectra of olive leaf aqueous extracts in terms of total polyphenols concentration, antioxidant activity, and extraction yield for each extraction procedure. PLS models were used to view which PCA scores are the best suited as input for ANN based on three output variables. ANN showed very good correlation of NIRs and all tested variables, especially in the case of total polyphenolic content (TPC). Therefore, ANN can be used for the prediction of total polyphenol concentrations, antioxidant activity, and extraction yield of plant extracts based on the NIR spectra.

Keywords:
NIR spectra, artificial neural networks, olive leaf extracts, conventional extraction, microwave-assisted extraction, microwave-ultrasound-assisted extraction

Introduction

Polyphenolic compounds play an important role as health-protecting factors. They have been shown to exhibit different properties, such as anti-allergenic, anti-inflammatory, anti-microbial, anti-thrombotic, anti-cancer, cardio-protective, vasodilatory and antioxidant properties. For the food industry, polyphenolic compounds have the potential to be utilized as preservatives in foods and are often valid alternatives to synthetic food additives.

The olive (Olea europaea L.) belongs to the Oleaceae family and is native to tropical and warm temperate regions of the world. The olive is considered a multipurpose crop with great yield potential. The tree is famous for its fruit, and is commercially significant in the Mediterranean countries as a source of oil. Both the fruit and the oil represent important components in the daily diet of a large part of the world’s population due to their health benefits. The olive oil industry generates a large amount of by-products, such as crude olive cake, vegetation water, twigs, and leaves. Olive leaves are also one of the by-products of olive grove farming; they accumulate during the harvesting of the olive fruit. It has been reported in the literature that olive leaves represent a source of polyphenolic compounds with a wide range of physiological properties, such as anti-HIV properties, anti-proliferative and apoptotic effects, lipid-lowering activity, etc. Many of these properties are due to the antioxidant activity, which can be determined by different mechanisms, such as free radical scavenging, electron or hydrogen atom donation, or metal cation chelation. Furthermore, the health benefits of the olive can also be attributed to the polyphenols present in its leaves.

In order to acquire the highest possible amount of polyphenols, efficient extraction techniques are required. Some of the methods used nowadays include solid-liquid extraction, ultrasound-assisted extraction, microwave-assisted extraction, supercritical fluid extraction, and high pressure or pressurized liquid extraction. Solid-liquid extraction under controlled experimental conditions is important for the reproducibility of the products and the preservation of their bioactivity.
Once obtained, plant extracts need to be further treated in order to identify and quantify their chemical composition. Over the years, many analytical methods (spectrophotometry, gas chromatography, high performance liquid chromatography or capillary electrophoresis methods) have been used in identification and quantification of plant extracts\textsuperscript{24,26}. Listed methods usually demand complex sample preparation; they are time-consuming and not environmentally friendly. In order to ensure product development and quality control, near infrared (NIR) spectroscopy has become an accepted method for the qualitative and quantitative analyses of plant extracts. NIR is described as a non-invasive rapid method, which requires almost no sample preparation, as well as allows the possibility of on-/inline measurements, determination of physical and chemical parameters simultaneously, and can be applied for a wide range of samples\textsuperscript{27}. As presented by Belsčak-Cvitanović \textit{et al.}\textsuperscript{28}, NIR spectroscopy has great potential in analysis of the polyphenolic composition of medicinal plant extracts. Analyses of the NIR spectra can be performed using mathematical models and multivariate analysis (chemometrics). Chemometrics utilizes various multivariate techniques, such as principal component regression (PCR) and partial least squares (PLS), in order to build models with the use of reference chemical data obtained through other analytical techniques\textsuperscript{29}. However, in the case of the high non-linearity, the mentioned techniques lead to substantial errors. According to Dou \textit{et al.}\textsuperscript{30}, application of artificial neural networks (ANNs) can be a good alternative. ANNs are self-adaptive and massively parallel machine-learning systems composed of layers of processing elements (neurons), and are used primarily for solving pattern recognition problems by building nonlinear models. The models can be used to generalize their conclusions and to predict patterns that have not previously been encountered\textsuperscript{31}. Some examples of the use of NIR spectroscopy combined with ANNs are: (i) quantification of wine compounds\textsuperscript{32}, (ii) determination of tea polyphenols\textsuperscript{33}, (iii) characterisation of olive fruit and olive oil online\textsuperscript{34}, and (iv) determination of polyphenolic compounds of red wines\textsuperscript{35}. Although there exists research describing that the application of near infrared spectroscopy for determination of polyphenols in olive oil and even for the control of the whole process of olive oil production\textsuperscript{36}, in this work, the olive leaf only was analysed. The main objective of this study was to evaluate the ability of artificial neural networks in near infrared spectra calibration models to predict the total polyphenolic content, antioxidant activity, and extraction yield of the olive leaf aqueous extracts, prepared with three extraction procedures (conventional extraction, microwave-assisted extraction, and microwave-ultrasound-assisted extraction). To our knowledge, there are no papers that correlate olive leaf extraction using water as a solvent in terms of NIR spectroscopy application for monitoring the total polyphenolic content, antioxidant activity, and extraction yield from olive leaves. Most of the research done on olive leaves is focused on extraction optimization for obtaining the highest possible polyphenolic content, as well as the use of different organic solvents, excluding water.

**Materials and methods**

**Materials**

**Plant materials**

Dried olive leaves (\textit{Olea europaea} L.) were purchased from a specialized herbal store (Suban d.o.o., Zagreb, Croatia). Plant material was collected in the southern part of Croatia during the season of 2015, dried, and properly stored until used. Dry matter content of the plant was determined gravimetrically by a standard AOAC method\textsuperscript{36}.

**Chemicals and reagents**

Folin–Ciocalteu reagent and sodium carbonate were purchased from Kemika (Zagreb, Croatia). Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) and potassium peroxodisulphate were obtained from Fluka (Buchs, Switzerland). DPPH (1,1-diphenyl-2-picrylhydrazyl) and gallic acid (3,4,5-trihydroxybenzoic acid) were obtained from Sigma-Aldrich Chemie (Steinheim, Germany), while methanol was obtained from J. T. Baker (Deventer, The Netherlands).

**Methods**

**Milling**

Dried olive leaves were milled using IKA Tube mill control (IKA-Werke, Staufen, Germany). Milling conditions were as follows: 15000 rpm with adjusted milling time ($t = 10 – 40$ s) in order to obtain different particle size fractions.

**Separation of particle size fractions**

Milled plant material was subjected to sieving in order to separate the particle size fractions. The following standardized DIN sieves (Fritsch, Idar-Oberstein, Germany) were used: $d = 100, 250, 355, 500, 800$, and $1000$ μm pores diameter. Sieves were placed on a laboratory shaker and shaking was performed for 10 minutes in intervals of 3 seconds with amplitude of 3 mm. Seven different particle size
fractions were obtained: <100 μm, 100 μm, 250 μm, 355 μm, 500 μm, 800 μm, and 1000 μm. Three fractions (d = 100 μm, 355 μm, 500 μm) were the subject of following experiments.

Conventional extraction of the polyphenolic compounds

An amount of m = 1 g of dry plant material was placed in a V = 200 mL glass with V = 50 mL of deionised water, covered with aluminium foil, and heated to a specific temperature (T = 40, 60, 80 °C ± 0.5 °C) using the Ika HBR4 digital oil-bath (IKA-Werk GmbH & Co. KG, Staufen, Germany). Extraction experiments were performed based on the Full Factorial Experimental design (Table 1). After finishing the extraction processes, samples were filtered through a 100 % cellulose paper filter (LLG Labware, Meckenheim, Germany) with d = 20 – 25 μm pore size, and stored at T = 4 °C until analysed. All experiments were carried out in duplicate, and the results are expressed as the average ± standard deviation.

Microwave and microwave-ultrasound-assisted extractions of the polyphenolic compounds from olive leaves were performed in the MW-ER-02 (Lab Kits, USA). Effect of the microwaves and simultaneous effect of the microwaves and ultrasound (P = 50 W, f = 40 kHz) on the extraction efficiency was investigated. An amount of m = 1 g of dry plant material was placed in an Erlenmeyer flask V = 250 mL with metal part for the conduction of the ultrasound waves with V = 50 mL of deionised water. The extractions were performed based on the Full Factorial Experimental design shown in Table 2.
ishing the extraction processes, samples were filtered through a 100 % cellulose paper filter (LLG Labware, Meckenheim, Germany) with \(d = 5 – 13\) \(\mu m\) pore size, and stored at \(T = 4^\circ C\) until analysed. All experiments were performed in duplicate, and the results are expressed as the average ± standard deviation.

The total polyphenolic content (TPC) was determined spectrophotometrically using the Folin-Ciocalteu reagent, according to Lachman et al.\(^3\). The results were expressed as mg of gallic acid equivalents (GAE) per gram of dry matter (DM) of plant material. Antioxidant activity (AA) was assessed by the DPPH free radical assay\(^9\). The results were expressed as Trolox equivalents per gram of dry matter (DM) of plant material. All experiments were carried out in duplicate, and the results are expressed as the average ± standard deviation.

**Extraction yield**

Extraction yield (EY) was expressed as the dry matter content of the aqueous extracts. Dry matter content of the plant aqueous extracts was determined gravimetrically by a standard AOAC method\(^3\).

**NIR spectroscopy**

Measurements of the spectra were performed using NIR spectrophotometer NIR128L-1.7 (Control Development, South Bend, Indiana, USA) with installed Control Development software Spec32 using a halogen light source (HL-2000) for the wavelength range of \(\lambda = 904 – 1699\) nm. Three consecutive runs for every sample of aqueous extract were initially recorded across the entire spectral range, and the average spectrum was used for analysis. No mechanical or chemical treatment of the samples was needed prior to NIRS measurements.

**Principal Component Analysis**

Principal component analysis (PCA) was used for identifying patterns and highlight similarities and differences in the data of the individual set of experiments. The goal of PCA is to extract the important information from the data table and express this information as a set of new orthogonal variables called principal components or factors. Raw spectra were used to perform principal component analysis (PCA) by the Unscrambler® X 10.4, software (CAMO software, Norway).

**Partial least squares (PLS) regression modelling**

To predict total polyphenolic content, antioxidant activity, and extraction yield of the prepared extracts based on NIR spectra, partial least squares (PLS) regression was applied. PLS prediction is achieved by extracting a set of orthogonal factors called latent variables that have the best predictive power from the predictors. PLS regression modelling was carried out on raw NIR spectra using Unscrambler® X 10.4, software (CAMO software, Norway).

**Artificial neural network modelling**

To predict total polyphenolic content, antioxidant activity, and extraction yield of the prepared extracts based on NIR spectra, artificial neural network modelling was applied. Multiple layer perceptron networks were developed in Statistica v.10.0 software (StatSoft, Tulsa, USA). Based on the PLS analysis, the most suitable PCA factors were selected and used as the input variables. Artificial neural network (ANN) training was performed with separation of data into training, test, and validation sets at a 70:20:10 ratio. Back error propagation algorithm available in Statistica v.10.0 was applied for the model training. The model performance was evaluated based on \(R^2\) and root mean squared error (RMSE) values for training, test, and validation.

**Results and discussion**

**NIR spectra of olive leaf aqueous extracts**

In order to test the ability of ANN calibration to predict the total polyphenols concentration, antioxidant activity, and extraction yield of the olive leaf aqueous extracts based on NIR spectra, the raw spectra of extracts obtained by three different extraction procedures were recorded. For all the experiments presented in Table 1 and Table 2, duplicate measurements were performed. From these duplicate measurements, average values of spectra were used for later processing. According to literature\(^17\), baseline drifts are eliminated and small spectra differences are enhanced by pre-processing of spectra in terms of calculating first and second derivatives. Also, an enhancement of noise as a consequence of differentiation is avoided by smoothing the spectra in advance\(^18,19\). Nevertheless, in this work, the focus was on the raw spectra only in order to simplify data processing. The example of raw spectra for classical extraction procedure is presented in Fig. 1.

NIR spectra of the olive leaf extracts show no significant differences in the range of 904 to 1699 nm, based on the absorbance. However, in the ranges 904–928 and 1350–1699 nm, slight shifts are visible, indicating changes in the third and second overtone of the C–H and O–H relations\(^20\). These relations are also related to the hydroxyl group (–OH) bound directly to an aromatic hydrocarbon group\(^28\).
ANN modelling

As Dou et al. explained, ANNs suffer from three major drawbacks: (1) the predictive properties of ANNs strongly depend on the learning parameters and the topology of the network; (2) the training time is long; and (3) ANNs models are complex and difficult to interpret. Since all the NIR spectra used in this work had 793 wavelengths, in order to reduce this vast amount of data to be used later for ANN modelling, principal component analysis was performed to obtain coordinates of factors (scores). In most studies, PCA was used to scale down NIRs wavelengths to match the column number of output variables that are used for ANN, thus shortening the training time of the neural networks.

Graphical representation of correlations between PCA factors in combination with four parameters of the extraction process (temperature, time, particle diameter and rpm) and TPC, AA, and EY for each type of extraction are presented in Fig. 2.

For the classical extraction (Fig. 2a), TPC and AA were represented in third quadrant, and all the factors that were in the third quadrant were taken as inputs for ANN calculations, while the TPC and AA...
were used as output variables. Extraction yield was represented in the second quadrant, and all the factors and process parameters from the second quadrant were taken as input for ANN calculations and extraction yield as output variable. For microwave-assisted and microwave-ultrasound-assisted extraction, the procedure was the same, except for the case of AA in microwaves and ultrasound-assisted extraction (Fig. 2c), where AA was represented between the second and the third quadrant. In that case, all the factors and process parameters that were in the second and third quadrants were taken as input variables, and AA as an output variable. Although there were 26 factors and four process parameters at the beginning in all cases, a reduction in the number of input variables for ANN was observed, especially in the case of classical extraction for extraction yield where only 6 variables were selected, and also in the case of TPC and AA where 8 variables were selected as inputs for ANN.

Artificial neural network (ANN) modelling was performed with the number of the neurons in the hidden layer in a range from 3 to 13, depending on the extraction procedure. Optimal neural networks were selected based on the linear correlation coefficient ($R^2$) and the values of the root mean square error (RMSE). The $R^2$ values should have been as close to 1 as possible. Thus, a good-fitting model would have $R^2$ values above 0.90. Values between 0.70 and 0.90 indicate that the models can be considered fairly precise, and values below 0.70 indicate that the model can be used for qualitative distinction without the capacity to be used in quantitative prediction21,22.

The results of ANN in terms of correlation coefficient ($R^2$) and the values of the root mean square error (RMSE) are presented in Table 3. The graphical presentation of training, test, and validation for classical extraction, microwave-assisted extraction, and microwave-ultrasound-assisted extraction are presented in Fig. 3.

Based on the results presented in Table 3, the best values in terms of correlation coefficients for validation were achieved for TPC for all three different extraction procedures, with values of $R^2$ ranging from 0.8986 (microwave-ultrasound-assisted extraction) to 0.9807 (microwave-assisted extraction). Such good correlations are visible in Figs. 3a, 3d and 3g. With such high correlations, it can be stated that NIRs can easily be used as one of the techniques for determination of TPC concentration, which is in accordance with literature data regarding NIRs and TPC23,24. The values of correlation coefficients for validation for AA were also very high for classical extraction (0.9481) and microwave-assisted extraction (0.9742). The lowest correlation was obtained for the microwave-ultrasound-assisted
extraction; $R^2 = 0.8840$. Although the measurements were performed in duplicate and mean values were used for calculations, an error occurred.

For the extraction yield in all three extraction types, validation correlation coefficients were the lowest in the range of 0.8768 for the microwave-ultrasound-assisted extraction to 0.8911 for the classical extraction, probably due to experimental method used for the measurement of the extraction yield. In Fig. 3c, suggests that the extraction yield measured for the extract prepared under the conditions $t = 5 \text{ min}$, $T = 80 \degree \text{C}$, rpm = 500, and $d = 355 \mu\text{m}$ shows the highest data dispersion. Based on all obtained results, it can be concluded that the ANN showed very good correlation of NIRs and all tested variables (total polyphenol concentrations, antioxidant activity, and extraction yield), especially in the case of TPC. In addition, developed ANN model ensured good fitting based on the error values. For all selected ANN models, the validation error was low and slightly higher than the training error. Therefore, ANN can be used for the prediction of total polyphenol concentrations, antioxidant activity, and extraction yield of plant extracts based on the NIR spectra.

**Conclusions**

It is possible to obtain very good correlations with total polyphenol concentrations using ANNs even from the raw NIR spectra of olive leaf extracts, without pre-processing and additional manipulation of spectral data in terms of smoothening or differentiation. The obtained results show the potential of quantitative prediction of polyphenols with the possibility of on-line monitoring of total polyphenol concentrations in extraction procedures. In addition, very good correlations were obtained for the antioxidant activity and extraction yield, showing growing potential of NIR spectroscopy related to ANN for use in quality control.
Literature


27. Cozzolino, D., Infrared spectroscopy as a versatile analytical tool for the quantitative determination of antioxidants in agricultural products, foods and plants, Antioxidants 4 (2015) 482. doi: https://doi.org/10.3390/antiox4030482