Response surface methodology and artificial neural network modeling as predictive tools for phenolic compounds recovery from olive pomace

Ana Rita Silva [a,b,c], Manuel Ayuso [a,b], Taofiq Oludemi [a,b,d], Alexandre Gonçalves [e], Bruno Melgar [a,*], Lillian Barros [a,b,*]

[a] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
[b] Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
[c] Departamento de Ciencias Farmacéuticas. Facultad de Farmacia, CIETUS-IBSAL, Universidad de Salamanca, 37007 Salamanca, España
[d] Universidade de Vigo, Nutrition and Bromatology Group, Department of Analytical Chemistry and Food Science, Faculty of Science, E-32004 Ourense, Spain
[e] Collaborative Laboratory Mountains of Research (MORE), Brigantia Ecopark, 5300-358, 16 Bragança, Portugal

ARTICLE INFO

ABSTRACT

This study optimized the extraction of three major phenolic compounds (oleuropein, tyrosol, and verbascoside) from olive pomace using microwave- and ultrasonic-assisted methods. Screening factorial design (SFD) and central composite design (CCD) were employed, and response surface methodology (RSM) and artificial neural networks (ANN) were used for data modeling. The microwave-assisted method in the SFD yielded higher compound amounts, with verbascoside showing a four-fold increase compared to the ultrasonic-assisted method. Factors like vessel diameter, ultrasonic power using UAE, and solvent acidity in both techniques had minimally impacted extractability. CCD-RSM revealed temperaturés significantly affect on oleuropein, but improved tyrosol recovery, with the effect on verbascoside being influenced by the temperature range. RSM and ANN integration enhanced understanding and prediction of factor behavior. Microwave-assisted extraction at 113 °C for 26 min, with minimum ramp time of 7.7 min, yielded 67.4, 57, and 5.1 mg of oleuropein, tyrosol, and verbascoside per gram of extract, respectively, with a prediction error ranging from 0.83 to 15.19.

1. Introduction

Olive oil production is mainly concentrated in Europe (76 %), with Spain as the leading producer, followed by Italy, Greece, and Portugal. Despite the economic importance of this food product, the olive oil industry is responsible for a significant environmental footprint, mainly due to residue generation [30]. Olive pomace (OP) is the primarily bio-residue of olive oil production (approximately 35–40 kg per 100 kg of olive) and presents a substantial environmental problem with high cost of transportation and treatment for the industry [20,24]. Over the last few years, olive mills have transitioned from a conventional three-phase system to a more environmentally friendly two-phase system. This technological change improved the olive oil quality and reduced wastewater generation. Nevertheless, OP management still present a great challenge to the industry due to its high phytotoxicity. Thus, over 400 000 tons of olive pomace are produced yearly, harming the environment when not adequately treated [21].

This bio-residue is a source of high-value-added biomolecules such as lignin, cellulose, hemicellulose, minerals, organic acids, and phenolic compounds [27]. The wide structural diversity of these extractable bioactive molecules is one of the crucial factors for their exploitation by the cosmetic, food, and pharmaceutical industries. The most abundant phenolic compounds in OP are oleuropein, hydroxytyrosol, verbascoside, and tyrosol, well known for their ultraviolet radiation protection, antioxidant, anti-inflammatory, anti-melanogenic, cellular senescence delay, and other characteristics related to photoprotection and suppression of both intrinsic and extrinsic ageing signs [11,15,16,18,19,25,28,4,32], making them promising cosmetic ingredient.

Therefore, the recovery of these value-added compounds from OP and their subsequent conversion into products for different markets falls within the bioeconomy and circular economy paradigms. Novel extraction methods, including ultrasounds (UAE) and microwave-

assisted (MAE) extractions, have been utilised as efficient techniques to improve the recovery of phenolic compounds from different biological raw materials, offering advantages such as faster heat transfer, low solvent consumption, reduced extraction times, and higher extraction yields [29].

The challenges to improving the extractability of these compounds in olive pomace are related to the need to utilise olive pomace in its original state (over 70 % moisture content). Currently, the recovery of bioactive compounds from olive pomace is conducted using its dried form [2,5,26]. However, this process requires high energy and capital investment to effectively dry OP and enhance the extractability of its bioactive molecules. In addition, individual screening of extraction variables is a time-consuming process that ignores a possible interaction between factors. Thus, establishing a statistical design of experiment (DoE) approach for optimal variable combinations is required to maximize the final response [7]. In this context, the application of mathematical models, such as Response Surface Methodology (RSM) and Artificial Neural Networks (ANN), are useful to optimise the interactions between different variables assessed to guarantee optimum yields.

Therefore, the present work aims to optimise olive pomace extraction in its original form using two green extractions approaches (microwave and ultrasound assisted extraction). A Central Composite Design (CCD) was applied after a step-by-step screening analysis to assess the linear, quadratic, and interaction effects of the independent variables (time, temperature, solvent proportion, solid–liquid ratio, ramp, and power) on the target responses (verbascoside, tyrosol, and oleuropein), fitting the model through RSM and ANN. Hence, the extrapolation of the knowledge collected during the application of these two extraction technologies will be useful in the development of an extraction pilot system that will ensure optimum extraction yield of these compounds and contribute to small and medium-scale olive mills economy.

2. Materials and methods

2.1. Chemicals and standards

High-Performance Liquid Chromatography (HPLC)-grade acetonitrile (99.9 %) was acquired from Fisher Scientific (Lisbon, Portugal) and formic acid from Panreac (Barcelona, Spain). Phenolic compound standards were purchased from Extrasynthese (Genay, France). All other chemicals were of analytical grade and purchased from common sources. Water was treated using a Milli-Q water purification system (TGI

Pure Water Systems, Greenville, SC, USA).

2.2. Sample preparation

Olive pomace was obtained from Trás-os-Montes Prime Lda, a local olive oil producing company in Suções, Mirandela, Northeast of Portugal (N41.4909, W7.2603), with a two-phase decanter for olive oil recovery. On reception, the OP was placed in separate air-tight bags and stored at −20 °C until analysis.

2.3. Extraction procedure

The OP was extracted in its original form using UAE and MAE following conditions mentioned below. Afterwards, to homogenise the samples and dilute them in the appropriate solvent, extract suspension was freeze dried prior to chromatographic analysis.

2.3.1. Ultrasound-Assisted Extraction (UAE)
The extraction was performed in an ultrasonic homogenizer (CY-500 model, Optic Ivymen System, Barcelona, Spain) according to the following four continuous factors and ranges: X_1 – Solvent (0 – 1.25 % of acetic acid), X_2 – Time (1 – 25 min), X_3 – Ratio (25 – 75 g/L), X_4 – Power (125 – 450 W) and the categorical: X_5 – vessel diameter factor (60 or 80 mm diameter; small or big, respectively).

2.3.2. Microwave-Assisted Extraction (MAE)
The extraction was performed in a microwave digestion system (Speedwave Xpert, Berghof, Eningen, Germany). As in the UAE, homogenization/nomenclature of the factor was chosen as close as possible. Therefore, 4 continuous factors (and ranges): X_1 – Solvent (0 – 1.25 % of acetic acid), X_2 – Time (5 – 25 min), X_3 – Ratio (25 – 75 g/L), X_6 – Temperature (80 – 121 °C), and the categorical: X_7 – ramp factor (1 or 10 min) were selected. After the screening analysis, the three most important factors were selected, employing X_2 – Time (10 – 30 min) with a slight ascend to the right, X_6 – Temperature in the same ranges, and converting the categorical X_7 – ramp (5 – 15 min) into a continuous factor, also with a slight ascend to the right. All the above-mentioned factors and levels for both extraction techniques were used for the screening analysis.

2.4. Chromatographic conditions

The phenolic fingerprint of the extracts was determined by High-

Performance Liquid Chromatography (HPLC) (Prominence CBM 20-A, Shimadzu, Japan) with a UV-DAD detector. Extracts were redissolved in 20 % aqueous ethanol at 10 mg/mL and filtered through an LC filter disk (nylon filter 0.2 μm, 25 mm diameter, Whatman M, GE Healthcare, Buckinghamshire, UK). The HPLC column was a Kinetex C18 XB-C18 (5 μm, 250 mm, 4.0 mm) and the detection was made at 280, 330, and 370 nm as preferred wavelengths following a procedure previously described [22]. Quantitative analysis was performed using selected calibration curves obtained from commercial standards, and the results were expressed in mg per g of extract.

2.5. Experimental design

The optimisation was developed using three steps: 1) selection of extraction parameters according to a preliminary study [22], where different extraction conditions and extraction solvents were investigated to determine the optimum conditions for polyphenols extraction of irradiated samples, 2) screening of significant parameters using Minimum-Run Resolution IV Screening Factorial Design (SFD), and 3) optimisation of the significant parameters using Central Composite Design (CCD). The factors with *p*-values lower than 0.05 were considered as significant and hierarchy factors were maintained. SFD and CCD were built using Design Expert software, version 11.1 (Stat-Ease, Inc. Minneapolis, MN, USA) and RStudio, version 2022.12.0 + 353.

2.5.1. Minimum-run resolution IV screening factorial design (SFD)

SFD was utilised to identify the significant parameters that can influence phenolic extraction. The non-significant parameters from the SFD were screened out and fixed at convenience. SFD was based on the first-order polynomial model:

$$Y_i = C_0 + \sum_{i=1}^{n} C_i X_i + \sum_{i=1}^{n-1} \sum_{j \neq 1}^{n} C_{ij} X_i X_j + \varepsilon \tag{1}$$

Y_i is the experimental response, X_i and X_j are the independent variables, C_0, C_i, and C_{ij} are the regression coefficients for the intercept, linear terms, and interactions, respectively. The SFD at two levels (high and low) was used for the screening of the main effect, 2-factor interactions, and curvature from the factors described in section 2.3. The factors and ranges are presented in Fig. 1, section A1. The design consists of 12 runs of the combinations of the independent factors at high (+) and low (-) levels, including 4 additional runs at the base/medium level (0), for a final 16-run experimental design. The following mathematical relationship was utilised to convert a real value (z_i) into a coded representation (x_i) within the context of a specific experimental design:
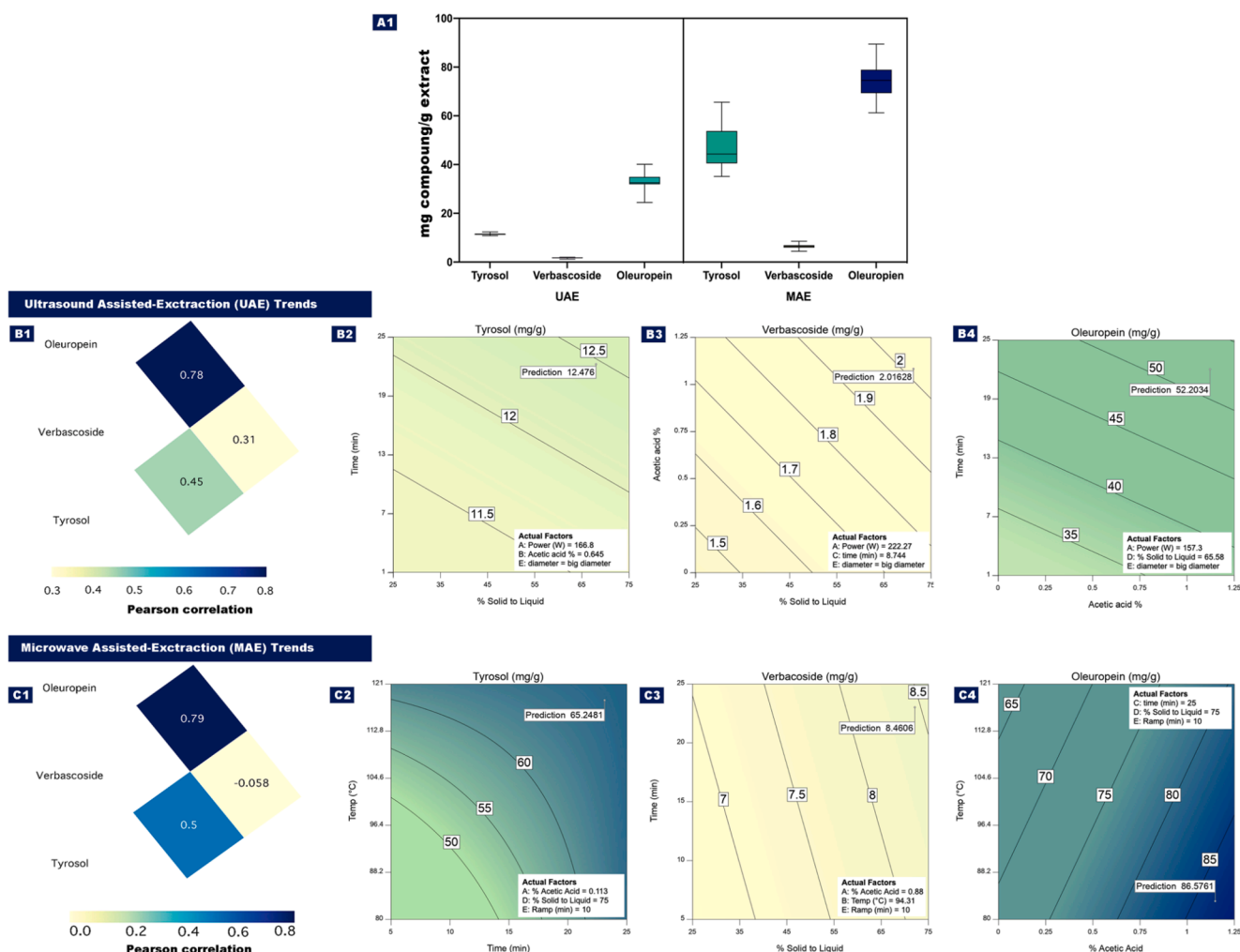
Fig. 1. A1) Descriptive statistics of the responses (tyrosol, verbascoside, and oleuropein) for UAE and MAE. Figures B clusters the information UAE while **Figures C** clusters MAE information. B1 and C1 displayed a diagonal Pearsońs correlation matrix of the compounds for each extraction, and figures B2-B4 and C2-C4) display in a 2D contour plots the minimum and maximum values recover, ascending directions and curvature from the extraction, presented separately for better comprehension of each extraction technology. On X-axis, the most significant factor is display while in the Y-axis the second most significant factor is presented and within the plot in a white dialog box, the rest of the factor are shown with the conditions selected for plotting.

$$x_i = \left(\frac{z_i - z_i^0}{\Delta z_i}\right)\beta_d \qquad (2)$$

Where Δz_i represents the distance between higher and lower real values to the central point of the variable, β_d is the higher code limit value in the matrix for each variable, and z^0 is the real value in the central point.

2.5.2. Central Composite Design (CCD)

CCD investigated the significant parameters obtained during SFD for further optimisation. Three factors were examined at five levels ($-\alpha$, -1, 0, 1, α) and the design was created in two blocks containing 6 centre points and 20 runs (Fig. 2 – A1). The data collected from the CCD was analysed using multiple linear regression to fit a quadratic polynomial model of the form:

$$Y = \beta_0 + \sum_{i=1}^{n}\beta_i X_i + \sum_{i=1}^{n}\beta_{ii}X_i^2 + \sum_{i=1}^{n-1}\sum_{j\neq1}^{n}\beta_{ij}X_iX_j + \varepsilon \qquad (3)$$

As in section 2.6.1, Y represents the predicted response, X_i and X_j represent the independent variables, β_0, β_i, β_{ii} and β_{ij} are the regression coefficients for the intercept, linear, quadratic, and interaction coefficients of the model, respectively.

2.5.3. Artificial neural network fitting

Artificial neural networks (ANNs) have been proposed as a promising method for generating non-linear models for response surfaces and optimisation within the field of analytical chemistry [33]. These networks are based on the structure of biological neural networks and comprise a collection of highly interconnected processing units known as neurons. Neurons are organised into a series of layers, including an input layer with neurons corresponding to independent variables, an output layer with neurons corresponding to dependent variables, and one or more intermediate layers referred to as hidden layers, which establish connections between inputs and outputs, through a process known as training [23].

Due to the limited number of data points collected, it became necessary to make specific adjustments in the R script (as outlined in the attached repository in the paper). To address this, data training and fine-tuning were executed using the 'neuralnet' function, with the specification of architecture type, error function, threshold, and fitting configurations. Subsequently, for the validation phase, we leveraged the optimization achieved through Response Surface Methodology (RSM). We provided the input data along with the experimental observations and compared them with the predictions generated by the neural network using the 'compute' function. These values were then subject to further evaluation, wherein we calculated metrics including Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R-squared).

The training was carried out using backpropagation by adjusting the connections between neurons to adapt the outputs to the desired values
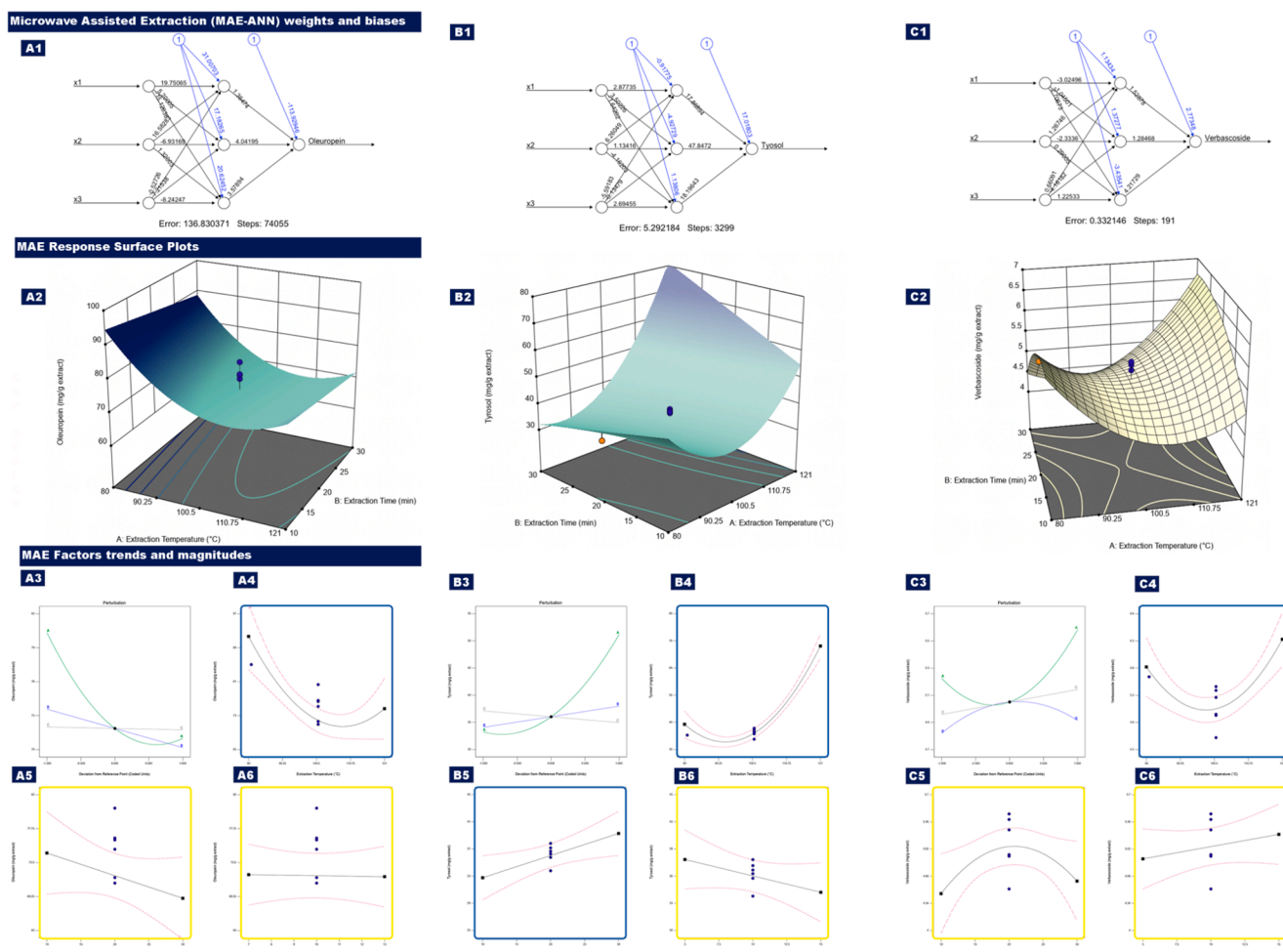
Fig. 2. Central Composite Design of MAE is displayed in different layer following the color scale mentioned in section 2.6.1. A1, B1, and C1 summarize the weights and biases of the ANN modeling while A2, B2 and C2 show the response surface graphs for each extracted compound. A3, B3 and C3 display the perturbation plot of the 3 factors analyzed. while A4-A6 display the single-factor behavior of oleuropein highlighting the border of the plot in blue for significant factors and yellow for not significant ones. Finally, B2-B6 and C2-C6 present the same order and meaning that A4-A6 but for tyrosol and verbascoside respectively.

of the dependent variables using a fixed parameter which in our case was targeted to minimise error. In the training phase, each number of neurons receives the input signal x_i from n neurons, aggregate them by using the adjusted weights (w_{ij}) of the joints, better known as synapses. The results are then passed through a selected function transformation which will then output the signal y_i, regarding the adjustment to the previously mentioned statistical fitting coefficients. The selected transformation functions tested were linear, sigmoid/logistic, hyperbolic tangent, soft plus, and ReLU using sum square error (SSE) as an error function with a threshold of 0.05 to find the best fitting [17].

2.5.4. Analysis, optimisation, and confirmation of prediction models

Both RSM and ANN-developed models were tested on their prediction performance, making comparisons and computing different fitting statistic coefficients (loss functions) such as the R^2, mean absolute error, RMSE, and mean absolute percentage error (MAPE) values. The last three were targeted for a minimum value and the first one targeted for maximum value for a well-fitted model:

$$Mean\,absolute\,error = \frac{1}{n}\sum_{i=1}^{n}|(p_i - t_i)| \qquad (4)$$

$$RSME = \sqrt[2]{\frac{\sum_{i=1}^{n}(p_i - t_i)^2}{2}} \qquad (5)$$

$$MAPE = \frac{1}{n}\sum\left|\frac{y_i - \widehat{y_i}}{y_i}\right| * 100 \qquad (6)$$

Where t_i is the value of the experimental data, p_i is the predicted value, and n is the number of samples in the dataset. Additionally, yi is the average of the experimental response, while $\hat{y}_i$ is the average of predicted response. The optimised network and RSM for each response were tested, and the results were plotted against true values. Finally, the global optimisation points were generated by Design expert software and the confirmation points were analysed in both RSM and ANN model prediction, comparing finally the Absolute Residual Error using the following equation:

$$Absolute\,Residual\,error = \left|\frac{y_i - \widehat{y_i}}{y_i}\right| * 100 \qquad (7)$$

2.6. Statistical analysis

Different tools were used for assessing and evaluating the data collected. Analysis of variance (ANOVA) for the determination of the individual linear, quadratic, and interaction regression coefficients was carried out using Design Expert software, version 11.1 (Stat-Ease, Inc. Minneapolis, MN, USA). Afterwards, the selection of the final model was achieved by employing the Bayesian information criterion (BIC) either in the forward or backward form, considering the following definition to compute BIC:

$$BIC = k\ln(n) - 2\ln(\widehat{L}) \qquad (8)$$

Where k stands for the number of parameters estimated by the model, n is the number of observations, and $\widehat{L}$ the maximised value of the likelihood function of the model. The final model was assessed by different criteria such as model significance (computing the F value at $p < 0.05$), the Δ of R^2_{adj} and $R^2_{pred} < 0.2$, and computing the normality of their residuals. ANN, Mean Absolute Percentage Error (MAPE), and Root Squared Mean Error (RMSE), and prioritising the factorial hierarchy and evaluating the pertinent fit statistics, the data was processed and analyzed using RStudio, version 2022.12.0 + 353 using the following libraries: Experimental design (DoE.base, SixSigma, AlgDesign, rsm, DoE.wrapper, AICcmodavg), Data wrangling and visualization (tidyverse, ggplot, ggpurb, multipanel figure, magrittr, broom), and ANN

modeling (neuralnet, MASS, Metrics).

2.6.1. Exploratory data analysis (EDA)

The YLGnBu-7 colour scale (Fig. S1) was used to homogenise the obtained data and allow for easy visualization, where yellowish colors represent the lowest values, greenish colors for medium values, and blueish colors for the higher values. In all the constructed models the colour scale was always used to highlight the *p*-values, where the yellow colour shows a not significant effect ($p > 0.05$), light green ($p \leq 0.05$), turquoise colour ($p \leq 0.01$), and blue ($p \leq 0.001$) showing different significant effects.

As a parallel analysis, a correlation was performed for all the responses analysed within every extraction type using the functions provided in the Scipy package (release 1.6.1) in Python 3.8.12 using the opensource IDE Spyder 5.1.5. The result heatmap visualization and EDA was achieved using Numpy (version 1.19.2), Pandas (version 1.1.3), matplotlib (version 3.5.1) and seaborn (version 0.11.1).

3. Results and discussion

3.1. Screening of significant parameters using SFD

Fig. 1-A1 compiled the results from the identified individual compounds (tyrosol, verbascoside, and oleuropein) of the 16 experimental runs in a box-plot graph from all the ranges of factors tested (complete numerical data in Table S1). The maximum extractability was shown in MAE for tyrosol, verbascoside, and oleuropein (65.58, 8.56, and 89.45 mg/g of extract, respectively). Tyrosol and oleuropein in MAE presented an average value that is 10 and 15 times fold higher than verbascoside, Finally, the same pattern of the compounds yields remains similar in UAE. The compounds identified matched with some of our previous works and with other authors [9,22,31].

The depicted correlations at Figure B1 and C1 represents a Pearsons correlation coefficient, which quantifies the similarity between compounds based on their extraction behavior under identical conditions. Values nearing 1 signify a high degree of similarity among compounds, while values approaching -1 indicate substantial dissimilarities. For instance, in UAE and MAE (Fig. 1-C1), a high correlation value between oleuropein and verbascoside (0.78 and 0.79, respectively) and non-correlational values between oleuropein and tyrosol (-0.058) was observed. Understanding the meaning of those correlations will help anticipate the difference in extractability when one factor changes its levels. For example, the modification of some factorial magnitudes in the case of oleuropein and verbascoside (which have a correlation of almost 80 %), is expected to affect the amplitude of the response of both compounds equally.

The models were constructed by executing, in most cases, the BIC backward factor selection, prioritising the factorial hierarchy, as shown in Table 1. The precision of the models were suitable considering the R^2 value (which in all cases was above 0.9) and the delta of the adjusted determination coefficients which was equal or lower than 0.2. Finally, at the bottom of the same chart, a polynomial equation is provided, and the remaining compounds were also constructed.

Fig. 1-B2-B4 presented 2-dimensional contour plots containing the recovered concentration of the compounds of interest (mg compound/g extract of olive pomace) through UAE technology. In the "*X*" axis, the most significant factor of each response was represented, while in the "*Y*" axis, the second most significant factor was also mapped (according to Table 1). The other factors were fixed at their highest point, selected after an internal optimisation approximation, and displayed in the white text boxes within the graphs (numerical data in supplementary Table S2). From the three contour plots, verbascoside represents the least quantity of recovered extracted phenolic compound with minimal and maximum concentrations of 0.82 mg/g and 2.03 mg/g, respectively. Oleuropein displayed the highest amount (40.13 mg/g) and data spread (15.67 mg/g). Finally, an average of 11.49 mg/g of tyrosol, making it

Table 1
SFD Model coefficients and fit statistics of the results collected (green and yellow terms showing significance and not significance, respectively, NC – not considered, NA – not available).

Source	Tyrosol		Verbascoside		Oleuropein	
	UAE	MAE	UAE	MAE	UAE	MAE
Intercept	11.63	46.55	1.71	6.6	34.53	73.14
X1:% Acetic Acid	-0.014	3.11	0.1475	0.9988	3.91	6.83
X2: Time	0.3565	4.66	0.1425	0.3813	4.61	1.65
X3: % Solid to Liquid	0.3	4.5	0.1625	0.58	1.79	1.49
X4 or X6: Power or Temp	-0.0325	4.63	0.0075	-0.0112	-0.8229	-2.65
X5 or X7: Vessel Diameter or Ramp	0.021	6.89	0.015	0.6875	1.66	2.38
X1X2	NC	-3.93	-0.1219	-0.2556	NC	1.86
X1X3	NC	1.62	NC	0.5056	NC	NC
X1X4 or X1X6	-0.089	-1.99	0.0769	NC	NC	NC
X1X5 or X1X7	NC	-4.13	NC	-0.2694	NC	NC
X2X4 or X2X6	-0.2135	NC	-0.1419	NC	-4.97	-1.42
X2X5 or X2X7	NC	NC	NC	NC	NC	-2.81
X3X4	NC	NC	NC	NC	-0.7729	NC
X4X5	-0.249	NC	-0.2625	NC	-5.33	NC
X1X2X6	NC	6.99	NC	0.4588	NC	NC
Model	0.0213	0.0142	0.0415	0.0135	0.0203	0.0166
Curvature	0.2022	0.0088	0.1309	0.1408	0.0434	0.8298
Lack of fit	0.9703	0.9935	0.8976	0.8545	0.8426	0.9891
R²	0.968	0.9862	0.968	0.9657	0.9528	0.9289
Adjusted R²	0.872	0.9401	0.872	0.8884	0.8584	0.8153
Predicted R²	NA	0.9445	NA	0.6827	NA	0.6585
Δadj-Pred R²	NA	-0.0044	NA	0.2057	NA	0.1568
Model selection	BIC backward	BIC backward	BIC backward	BIC backward	BIC backward	BIC forward
Polynomial eq example (Tyrosol-UAE)	11.63 - 0.014*X1 + 0.3565*X2 + 0.3*X3 - 0.0325*X4 + 0.021*X5 - 0.089*X1*X4 - 0.2135*X2*X4 - 0..249*X4*X5					

2.8-fold lower than oleuropein but 6.6-fold higher than verbascoside was obtained. Power and diameter factor were not significant parameters in all the responses; therefore, they could be fixed for convenience in other experimental steps which required UAE. It is worth noting that the capabilities of the experimental design will compute values outside of the experiment's records. However, extrapolation must be ignored or treated cautiously, pointing out that suggested trends could be broken easily outside the factorial boundaries tested.

As in the UAE, the graphs Fig. 1 **C2-C4** show the three compounds extracted using the microwave-assisted technology, where the highest extraction was found on oleuropein with 89.45 mg/g with a combination of factors of *X1*: 1.25 % of acetic acid, *X2*: 25 min, *X3*: 75 % solid to liquid ratio, *X4*: 80 °C and *X5*: 1 min ramp to achieve the desired temperature. It is also interesting to point at a small inflection point in the higher ranges of tyrosol which show the ascending lines to flex as the navigation in the model increases. This suggests a slow-down in extractability in the higher levels of the time and temperature factors, with average values of 65.58 mg/g. On the other hand, verbascoside was the compound with the lowest extractability [6.5 mg/g] when compared with mean values of tyrosol (7.2-fold lower) and oleuropein (11.4-fold lower), and with a maximum extraction value of verbascoside registered at 8.56 mg/g.

Finally, comparing UAE vs MAE compounds extractability, the computed values of MAE were higher in every single compound, which is believed to happen due to the usage of higher temperatures. The comparison of the 3 compounds in MAE were 5.3, 4.2, and 2.2-fold higher for tyrosol, verbascoside, and oleuropein, respectively, when comparing the maximum extractability values against UAE. Therefore, MAE offers at least double the amount of compounds extractability within the factorial levels tested. Other authors also reported better extractability at higher temperatures using the microwave-assisted extraction [10,12].

In summary, experimental data from the Screening Factorial Design has proved that in the conditions tested, microwave-assisted technology recovers a higher yield of the 3 compounds compared with ultrasonic-assisted extraction. The higher yield of the compounds was found on oleuropein, followed by tyrosol, and finally verbascoside. Factors such as vessel diameter, and ultrasonic power provided no or little

significance to the extractability of the compounds when employing UAE. Oleuropein and verbascoside are the higher correlated compounds extracted, while the behaviour within tyrosol and oleuropein seems to be independent.

3.2. Central composite design of the MAE

Based on the information extracted from the SFD, MAE was selected as the best extraction technology. Factors such as extraction temperature, time, and ramp time were selected as the main parameters to explore, while other factor tested were keep fix, compute, and optimise in the CCD. In addition, extraction time and ramp time were moved slightly on the increased values range, as shown in Table 2.

The CCD experimental design presented in Table 2 was performed in duplicate to generate a higher volume of data points and accuracy without requiring extensive repetitions. Therefore, data processing was treated as the means of both values of each run and the following computing performed in the same fashion.

The minimum recovered mean value in the CCD experimental runs was obtained for verbascoside (5.16 mg/g) followed by tyrosol (40.21 mg/g) which was 7.8 times fold-higher, and finally, oleuropein (74.89 mg/g) which presented around 14.5-, and 7.8-times fold-higher values compared to the extracted verbascoside and tyrosol, respectively Table 2.

3.2.1. RSM factorial behavior and optimum point

The normal process when modeling through RSM consists of selecting factors that provide significant differences while maintaining usual statistical diagnostics tuned to their better results, this process then becomes searching for better trade-offs, where single and global optimisation will be handled [8]. Fig. 2 **A2-A6, B2-B6, and C2-C6** showed a graphical representation of the RSM modeling for oleuropein, tyrosol, and verbascoside, respectively. **A2, B2,** and **C2** displayed the final surface modeled, followed by the perturbation plot of the factors analysed in Fig. 2 **A3, B3, and C3**, which is a graphical representation used to assess the sensitivity of a response variable to changes in the factors or input variables when making small changes or perturbations to the factor levels. Additionally to the holistic view of the perturbations plots,

Table 2

CCD Experimental Design, RSM, and ANN modelled compound responses.

Run	Type	Temperature (°C)	Extraction Time (min)	Ramp Time (min)	Oleuropein (mg/g extract)			Tyrosol (mg/g extract)			Verbascoside (mg/g extract)		
		(Coded Values within)			Exp	RSM	ANN	Exp	RSM	ANN	Exp	RSM	ANN
1	Factorial	89 (-1)	14 (-1)	7 (-1)	86.22	84.48	84.58	34.91	35.76	34.63	5.46	5.82	5.20
2	Factorial	113 (1)	26 (1)	13 (1)	69.65	71.49	69.70	54.80	55.42	54.81	6.42	6.17	6.32
3	Central	101 (0)	20 (0)	10 (0)	80.32	74.01	71.47	37.82	35.84	36.11	5.26	5.03	5.10
4	Factorial	113 (1)	14 (-1)	13 (1)	79.72	82.30	79.68	38.21	44.36	38.21	5.46	5.60	5.42
5	Factorial	89 (-1)	26 (1)	13 (1)	74.71	73.07	77.61	32.10	30.36	34.66	4.24	4.23	4.10
6	Central	101 (0)	20 (0)	10 (0)	71.63	74.01	71.47	36.27	35.84	36.11	4.52	5.03	5.10
7	Factorial	113 (1)	14 (-1)	7 (-1)	62.12	67.38	65.37	48.10	46.53	48.12	4.46	4.58	4.49
8	Central	101 (0)	20 (0)	10 (0)	75.17	74.01	71.47	35.76	35.84	36.11	4.93	5.03	5.10
9	Factorial	89 (-1)	14 (-1)	13 (1)	85.55	83.88	82.92	35.56	33.59	34.58	5.12	4.80	5.13
10	Factorial	89 (-1)	26 (1)	7 (-1)	87.44	88.48	87.41	35.23	32.53	34.66	5.49	5.24	5.30
11	Factorial	113 (1)	26 (1)	7 (-1)	69.63	71.39	69.62	56.95	57.59	56.95	4.96	5.16	5.27
12	Central	101 (0)	20 (0)	10 (0)	76.34	74.01	71.47	33.79	35.84	36.11	5.39	5.03	5.10
13	Star	121 (1.682)	20 (0)	10 (0)	78.53	72.62	78.44	72.18	68.75	72.18	6.50	6.37	6.52
14	Star	101 (0)	20 (0)	5 (-1.682)	69.65	69.94	72.86	35.85	38.04	36.19	4.83	5.07	5.29
15	Central	101 (0)	20 (0)	10 (0)	70.94	69.73	71.47	37.16	36.21	36.11	4.95	5.07	5.10
16	Star	80 (-1.682)	20 (0)	10 (0)	85.04	88.32	87.72	35.29	38.62	34.61	5.64	5.77	5.23
17	Central	101 (0)	20 (0)	10 (0)	76.58	69.73	71.47	36.67	36.21	36.11	5.46	5.07	5.10
18	Star	101 (0)	10 (-1.682)	10 (0)	69.51	72.59	75.95	34.59	32.92	34.67	4.54	4.55	4.52
19	Star	101 (0)	30 (1.682)	10 (0)	62.94	66.87	67.01	37.22	39.50	37.24	4.56	4.55	4.66
20	Star	101 (0)	20 (0)	15 (1.682)	66.12	69.52	70.11	35.66	34.38	35.98	5.06	5.07	5.16

the Fig. 2 subfigures **ABC 4-6** revealed the importance of extraction temperature (significant factors) for the compounds analysed, where oleuropein presented thermolabile behavior, while tyrosol and verbascoside benefit from higher extraction temperatures. The second relevant factor was extraction time, but only tyrosol showed statistical significance. In fact, the factorial graph reveals an increase of tyrosol yield with higher temperature and extraction time. The third factor analyzed (Ramp time), did not shown statistical significance and no interaction of factors where spotted. The thermolabile behaviour of oleuropein has already been observed in other research in both MAE and UAE [1,14]. [3], using a high temperature and pressure reactor to extract compounds of interest from the solid residue of Italian black olives, found a similar behaviour to that described in our data for oleuropein and for the positive effect of temperature and extraction time on tyrosol. Regarding Verbascoside, other research using MAE technology showed that its extraction improves with a temperature increase, with time not being a significant factor in the extraction of the compound [13].

Summarizing the last paragraph, the conditions for maximizing the recovery yield must consider different factorial parameters, which are described in **Table S4**. To optimise the recovery of all compounds, global maximisation was used for all factors except ramp time which was not statistical significance and was minimised. Thus, the optimised conditions were as follows: extraction temperature equal to 113 °C, extraction time of 26 min, and a ramp of 7.7 min, to achieve a predicted 69.3, 57.5 and 5.3 mg /g of extract of oleuropein, tyrosol, and verbascoside, respectively (**Table S4**). Although the chosen conditions provided by the software for an overall optimisation are multiple and based on our experimental data and requirement, the higher temperature was selected to extract the compounds. In addition, lower ramp time was selected aiming to reduce energy and time consumption.

3.2.2. ANN and RSM models fitting

To achieve optimal modeling conditions, we rely on the visual representation of points in Fig. 3 together with a tool kit of fitting statistical computations like the one mentioned in section 2.8.

Regarding the visual representation, the **sections A1-B3** presented

the experimental values vs the predicted ones for every compound (Figure A1-A3), and the representation of all data (experimental, ANN and RSM predictions) for the yield of every compound against the experimental runs (Figure B1-B3). The overall visual representation shows trends and similarities which help to better understand the numerical coefficients reported on Fig. 3 C1 for RSM and Fig. 3 **C2** for ANN.

Fig. 3 **C1** display the compute coefficients for every single factor, interaction, and quadratic term in numerical values with their confidence interval (95 %). The colors scale was used to represent the coefficients statistical significance. The model fitting terms and the polynomial equation employed for the calculations were also presented in the lower part of this figure.

Regarding ANN coefficients, the **section C2** presented the neurons weight values used in the graphical plot displayed in Fig. 2 **A1, B1 and C1** for oleuropein, tyrosol and verbascoside respectively. In addition, this figure also showed the computed error and steps to allow the fitting of the model. All the numerical data from the ANN neurons and biases will help to interpolate values in the confirmation tests.

Detailed information about the model's efficiency is presented in Fig. 3 **C4,** and their robustness is provided in Fig. 3 **C3**. It is important to highlight that a good model fit is determined by high R^2 values and low RMSE and MAPE. Therefore, when comparing both models, oleuropein was a better fit (highest R^2) according to the RSM technique, while tyrosol and verbascoside were better fitted on the ANN model (Fig. 3 C3). At the same time, when testing the confirmation points, oleuropein and tyrosol showed better performance (lower absolute error) with the ANN model, while verbascoside was better fitted with the RSM model.

Although the efficiency of the neural networks improves with bigger training data, combining ANN with RSM provides better prediction values with less errors improving the final model, even with small dataset (Fig. 3 **C3-4**). Tyrosol and verbascoside were the response with higher absolute error on the confirmation point, and when comparing the real vs predicted values, a narrowed distribution is shown, which is corroborated by the higher R^2. The absolute errors on the prediction ranged between 0.83 and 15.19, with all the responses within the confidence interval of prediction. This behaviour might suggest a slight
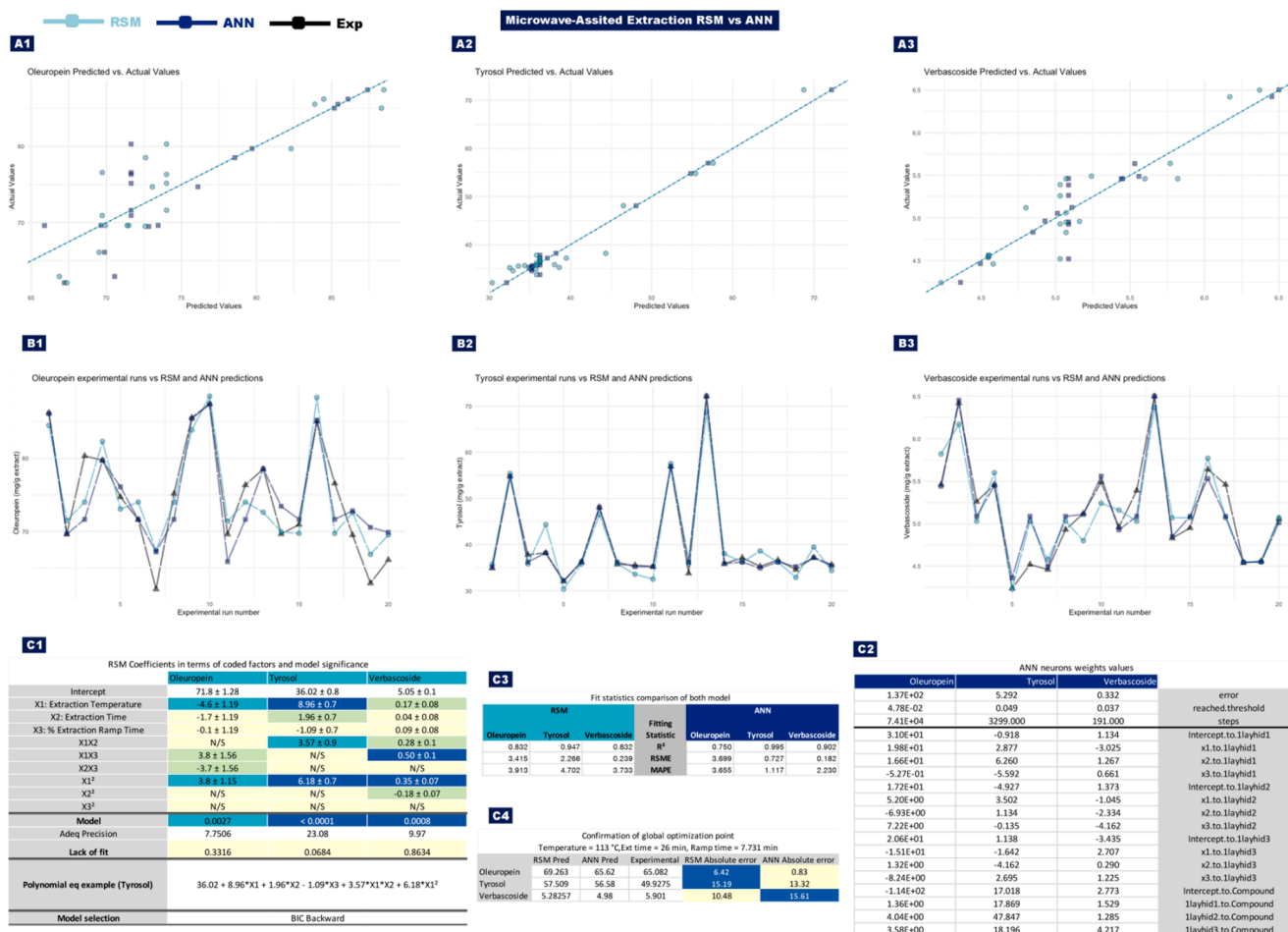
Fig. 3. Divided into three rows, the first row (A1-3) presents plots comparing the experimental values (black) with the predicted values computed by RSM (light green) and ANN (blue). The second row (B1-3) displays chronological plots of the experimental and predicted values against the experimental run order. In the bottom row, the figure includes tables of coefficients for RSM (labeled as C1) and ANN (labeled as C2). In the RSM coefficients table, factors that are not significant are highlighted in yellow, while factors with higher significance are represented by shades of green to blueish colors. Additionally, in the middle of the bottom row, table C3 presents the fit statistics computed for the model. Finally, table C4 shows the confirmation points along with their respective errors.

overfitting of data and refining the modelling could help to reduce the error, although the values presented are within the confidence interval of prediction.

Therefore, in the experimental tests performed and analysed, the ANN presented a slight edge over the RSM technique according to the fit statistics shown, allowing a better overall understanding of the behaviour of the factors and responses, in one hand, the visual aspect of RSM allow to understand trends and magnitude effect, while in the other hand, ANN allow another fitting approach beyond linear or quadratic modeling. This strategy helped to predict values within the ranges tested minimizing experimental runs. In other work, authors have also attributed a synergistic effect on the RSM-ANN modelling [6] as a hybrid optimisation approach.

4. Conclusions

This research successfully optimized the extraction of three bioactive compounds (tyrosol, oleuropein, and verbascoside) from pomace, a major by-product of the olive oil industry. This optimization approach addresses the environmental challenge posed by the accumulation of this waste by adopting a circular economy framework. The findings of this study contribute to the conversion of this waste into valuable resources.

The optimization process employed a combination of response surface methodology (RSM) and artificial neural network (ANN) modeling,

resulting in improved model fitting and prediction accuracy. The ultimate objective of this research is to apply the optimized parameters in a pilot extraction system, thereby facilitating the successful scalability of the process.

The optimized conditions determined through factorial combination involved using a microwave-assisted extraction at 113 °C for 26 min, with a minimum ramp time of 7.7 min. These conditions yielded predicted values of 67.4 mg/g for oleuropein, 57 mg/g for tyrosol, and 5.1 mg/g for verbascoside, respectively, per gram of extract. These results highlight the potential of the optimized extraction process in achieving high yields of the targeted bioactive compounds.

CRediT authorship contribution statement

Ana Rita Silva: Investigation, Writing – original draft, Conceptualization. **Manuel Ayuso:** Investigation, Writing – original draft, Conceptualization. **Taofiq Oludemi:** Investigation, Writing – original draft, Conceptualization. **Alexandre Gonçalves:** Writing – review & editing, Funding acquisition. **Bruno Melgar:** Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Visualization, Supervision, Project administration. **Lillian Barros:** Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

https://github.com/Bruno-melgar/Olea.

Appendix A. Supplementary data Repository

Supplementary data to this article can be found online at https://doi.org/10.1016/j.seppur.2023.125351.

References

[1] S. Achat, V. Tomao, K. Madani, M. Chibane, M. Elmaataoui, O. Dangles, F. Chemat, Direct enrichment of olive oil in oleuropein by ultrasound-assisted maceration at laboratory and pilot plant scale, Ultrason. Sonochem. 19 (4) (2012) 777–786, https://doi.org/10.1016/j.ultsonch.2011.12.006.

[2] P. Albahari, M. Jug, K. Radić, S. Jurmanović, M. Brnčić, S.R. Brnčić, D. Vitali Čepo, Characterization of olive pomace extract obtained by cyclodextrin-enhanced pulsed ultrasound-assisted extraction, LWT Food Sci. Technol. 92 (February) (2018) 22–31, https://doi.org/10.1016/j.lwt.2018.02.011.

[3] B. Aliakbarian, A.A. Casazza, P. Perego, Valorization of olive oil solid waste using high pressure-high temperature reactor, Food Chem. 128 (3) (2011) 704–710, https://doi.org/10.1016/j.foodchem.2011.03.092.

[4] M. Antónia Nunes, A.S.G. Costa, S. Bessada, J. Santos, H. Puga, R.C. Alves, V. Freitas, M.B.P.P. Oliveira, Olive pomace as a valuable source of bioactive compounds: A study regarding its lipid- and water-soluble components, Sci. Total Environ. 644 (2018) 229–236, https://doi.org/10.1016/j.scitotenv.2018.06.350.

[5] M. Antónia Nunes, S. Pawlowski, A.S.G. Costa, R.C. Alves, M.B.P.P. Oliveira, S. Velizarov, Valorization of olive pomace by a green integrated approach applying sustainable extraction and membrane-assisted concentration, Sci. Total Environ. 652 (2019) 40–47, https://doi.org/10.1016/j.scitotenv.2018.10.204.

[6] T. Aung, S.J. Kim, J.B. Eun, A hybrid RSM-ANN-GA approach on optimization of extraction conditions for bioactive component-rich laver (Porphyra dentata) extract, Food Chem. 366 (July 2021) (2022), 130689, https://doi.org/10.1016/j.foodchem.2021.130689.

[7] T. Belwal, P. Dhyani, I.D. Bhatt, R.S. Rawal, V. Pande, Optimization extraction conditions for improving phenolic content and antioxidant activity in Berberis asiatica fruits using response surface methodology (RSM), Food Chem. 207 (2016) 115–124, https://doi.org/10.1016/j.foodchem.2016.03.081.

[8] M.A. Bezerra, R.E. Santelli, E.P. Oliveira, L.S. Villar, L.A. Escaleira, Response surface methodology (RSM) as a tool for optimization in analytical chemistry, Talanta 76 (5) (2008) 965–977, https://doi.org/10.1016/j.talanta.2008.05.019.

[9] M.J. Cejudo Bastante, M. Chaalal, H. Louaileche, J. Parrado, F.J. Heredia, Betalain Profile, Phenolic Content, and Color Characterization of Different Parts and Varieties of Opuntia ficus-indica, J. Agric. Food Chem. 62 (2014) 8491–8499, https://doi.org/10.1021/jf502465g.

[10] G.S. da Rosa, S.K. Vanga, Y. Gariepy, V. Raghavan, Comparison of microwave, ultrasonic and conventional techniques for extraction of bioactive compounds from olive leaves (Olea europaea L.), Innov. Food Sci. Emerg. Technol. 58 (September) (2019), 102234.

[11] A.C.P. da Silva, J.P. Paiva, R.R. Diniz, V.M. dos Anjos, A.B.S.M. Silva, A.V. Pinto, E. P. dos Santos, A.C. Leitão, L.M. Cabral, C.R. Rodrigues, M. de Pádula, B.A.M. C. Santos, Photoprotection assessment of olive (Olea europaea L.) leaves extract standardized to oleuropein: In vitro and in silico approach for improved sunscreens, J. Photochem. Photobiol. B Biol. 193 (2019) 162–171, https://doi.org/10.1016/j.jphotobiol.2019.03.003.

[12] P. Darvishzadeh, V. Orsat, Microwave-assisted extraction of antioxidant compounds from Russian olive leaves and flowers: Optimization, HPLC characterization and comparison with other methods, J. Appl. Res. Med. Aromat. Plants 27 (August 2021) (2022), 100368, https://doi.org/10.1016/j.jarmap.2021.100368.

[13] A. Dobrinčić, M. Repajic, I. Elez Garofulić, L. Tuđen, V. Dragović-Uzelac, B. Levaj, Comparison of Different Extraction Methods for the, Processes 8 (9) (2020) 1008.

[14] M.N.M. Esther, C.T. Cristina, O. José, S.M. Rosario, G.L. Alonso, Oleuropein degradation kinetics in olive leaf and its aqueous extracts, Antioxidants 10 (12) (2021), https://doi.org/10.3390/antiox10121963.

[15] C.M. Galanakis, P. Tsatalas, I.M. Galanakis, Implementation of phenols recovered from olive mill wastewater as UV booster in cosmetics, Ind. Crop. Prod. 111 (April 2017) (2018) 30–37, https://doi.org/10.1016/j.indcrop.2017.09.058.

[16] W. Guo, Y. An, L. Jiang, C. Geng, L. Zhong, The protective effects of hydroxytyrosol against UVB-induced DNA damage in HaCaT cells, Phytother. Res. 24 (3) (2010) 352–359, https://doi.org/10.1002/ptr.2943.

[17] B. Jain, R. Rawat, QSAR and ANN-based molecular modeling, Comput. Modell. Simul. Designing Corros. Inhibitors 183–199 (2023), https://doi.org/10.1016/B978-0-323-95161-6.00006-0.

[18] S. Jeon, M. Choi, Anti-inflammatory and anti-aging effects of hydroxytyrosol on human dermal fibroblasts (HDFs), Biomed. Dermatol. 2 (1) (2018) 21, https://doi.org/10.1186/s41702-018-0031-x.

[19] S. Lammi, A. Barakat, C. Mayer-Laigle, D. Djenane, N. Gontard, H. Angellier-Coussy, Dry fractionation of olive pomace as a sustainable process to produce fillers for biocomposites, Powder Technol. 326 (2018) 44–53, https://doi.org/10.1016/j.powtec.2017.11.060.

[20] P. Leite, J.M. Salgado, A. Venâncio, J.M. Domínguez, I. Belo, Ultrasounds pretreatment of olive pomace to improve xylanase and cellulase production by solid-state fermentation, Bioresour. Technol. 214 (2016) 737–746, https://doi.org/10.1016/j.biortech.2016.05.028.

[21] J. Madureira, M.I. Dias, J. Pinela, R.C. Calhelha, L. Barros, C. Santos-Buelga, F.M. A. Margaça, I.C.F.R. Ferreira, S. Cabo Verde, The use of gamma radiation for extractability improvement of bioactive compounds in olive oil wastes, Sci. Total Environ. 727 (2020), 138706, https://doi.org/10.1016/j.scitotenv.2020.138706.

[22] J. Madureira, B. Melgar, C. Santos-Buelga, F.M.A. Margaça, I.C.F.R. Ferreira, L. Barros, S. Cabo Verde, Phenolic Compounds from Irradiated Olive Wastes: Optimization of the Heat-Assisted Extraction Using Response Surface Methodology, Chemosensors 9 (8) (2021) 231, https://doi.org/10.3390/chemosensors9080231.

[23] J. Nayak, K. Vakula, P. Dinesh, B. Naik, D. Pelusi, Intelligent food processing: Journey from artificial neural network to deep learning, Comput. Sci. Rev. 38 (2020), 100297, https://doi.org/10.1016/j.cosrev.2020.100297.

[24] M.A. Nunes, F.B. Pimentel, A.S.G. Costa, R.C. Alves, M.B.P.P. Oliveira, Olive by-products for functional and food applications: Challenging opportunities to face environmental constraints, Innov. Food Sci. Emerg. Technol. 35 (2016) 139–148, https://doi.org/10.1016/j.ifset.2016.04.016.

[25] S.H. Omar, Oleuropein in olive and its pharmacological effects, Sci. Pharm. 78 (2) (2010) 133–154, https://doi.org/10.3797/scipharm.0912-18.

[26] M. Panić, M. Radić Stojković, K. Kraljić, D. Škevin, I. Radojčić Redovniković, V. Gaurina Srček, K. Radošević, Ready-to-use green polyphenolic extracts from food by-products, Food Chem. 283 (October 2018) (2019) 628–636, https://doi.org/10.1016/j.foodchem.2019.01.061.

[27] N. Rahmanian, S.M. Jafari, C.M. Galanakis, Recovery and removal of phenolic compounds from olive mill wastewater, JAOCS, J. Am. Oil Chemists' Soc. 91 (1) (2014) 1–18, https://doi.org/10.1007/s11746-013-2350-9.

[28] F. Rodrigues, F.B. Pimentel, M.B.P.P. Oliveira, Olive by-products: Challenge application in cosmetic industry, Ind. Crop. Prod. 70 (2015) 116–124, https://doi.org/10.1016/j.indcrop.2015.03.027.

[29] A.R. Silva, J. Pinela, P.A. García, I.C.F.R. Ferreira, L. Barros, Cytinus hypocistis (L.) L.: Optimised heat/ultrasound-assisted extraction of tannins by response surface methodology, Sep. Purif. Technol. 276 (2021) 119358, https://doi.org/10.1016/j.seppur.2021.119358.

[30] T. Stillitano, G. Falcone, A.I. De Luca, A. Piga, P. Conte, A. Strano, G. Gulisano, A life cycle perspective to assess the environmental and economic impacts of innovative technologies in extra virgin olive oil extraction, Foods 8 (6) (2019), https://doi.org/10.3390/foods8060209.

[31] Y. Wang, G. Luan, W. Zhou, J. Meng, H. Wang, N. Hu, Y. Suo, Subcritical water extraction, UPLC-Triple-TOF/MS analysis and antioxidant activity of anthocyanins from Lycium ruthenicum Murr, Food Chem. 249 (January) (2018) 119–126, https://doi.org/10.1016/j.foodchem.2017.12.078.

[32] P. Xie, L. Huang, C. Zhang, Y. Deng, X. Wang, J. Cheng, Enhanced extraction of hydroxytyrosol, maslinic acid and oleanolic acid from olive pomace: Process parameters, kinetics and thermodynamics, and greenness assessment, Food Chem. 276 (16) (2019) 662–674, https://doi.org/10.1016/j.foodchem.2018.10.079.

[33] A. Xu, H. Chang, Y. Xu, R. Li, X. Li, Y. Zhao, Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review, Waste Manag. 124 (2021) 385–402, https://doi.org/10.1016/j.wasman.2021.02.029.