Food Control 184 (2026) 112011

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

HSI and ML for non-destructive pistachio quality assessment: Influence of location and irrigation on nutrient and fat composition

Raquel Martínez-Peña [a,b,1], Salvador Castillo [c,1], Sergio Vélez [d], Sara Álvarez [a,*]

[a] Instituto Tecnológico Agrario de Castilla y León (ITACyL), Ctra. Burgos km 119, 47071, Valladolid, Spain
[b] Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal (IRIAF), Centro de Investigación Agroambiental "El Chaparrillo" (CIAG), Ctra. de Porzuna s/n, 13071, Ciudad Real, Spain
[c] Cranfield Environment Centre, Cranfield University, Bedfordshire, UK
[d] JRU Drone Technology, Department of Architectural Constructions and I.C.T., University of Burgos, Burgos, 09001, Spain

ARTICLE INFO

Keywords:
Hyperspectral imaging (HSI)
Machine learning
Pistachio quality assessment
Predictive modeling
Non-destructive analysis

ABSTRACT

Pistachio quality is a major determinant of market value, particularly in Mediterranean environments strongly influenced by irrigation and climate. This study evaluated effects of orchard location and irrigation on pistachio nutritional content and assessed VIS–NIR hyperspectral imaging coupled with machine-learning techniques as a non-destructive predictive approach. Field experiments were conducted in 2022 in two commercial orchards in Castilla y León, Spain, under control and high-irrigation treatments. A total of 2818 pistachio kernels were analysed using a portable VIS–NIR hyperspectral camera (400–1000 nm). Mean spectra were extracted for each nut, pre-processed using standard normal variate correction, and linked to reference measurements of minerals, proximate composition and fatty acids. Partial Least Squares regression, Support Vector Regression and Extra Trees Regressor models were calibrated and validated to predict quality parameters from spectral data. Location, irrigation and their interaction significantly affected most nutritional and lipid traits. Pistachios from Moraleja de las Panaderas showed higher nitrogen, phosphorus, protein, ash and oleic acid contents, whereas samples from La Seca exhibited relatively higher sodium and linoleic acid levels. Increased irrigation enhanced the accumulation of several minerals and saturated fatty acids. Among the evaluated algorithms, Partial Least Squares regression provided the most consistent performance, accurately predicting nitrogen ($R^2 = 0.75$), zinc ($R^2 = 0.81$), oleic acid ($R^2 = 0.91$), linoleic acid ($R^2 = 0.87$), ash ($R^2 = 0.81$), carbohydrates ($R^2 = 0.87$) and humidity ($R^2 = 0.84$). Overall, VIS–NIR HSI with machine learning enables non-destructive, data-driven optimisation of pistachio irrigation management.

1. Introduction

Pistachio quality is a key determinant of market value, making it a central focus in orchard management and research (Khadivi et al., 2025; Martínez-Peña et al., 2023). In Mediterranean regions, where pistachios are a high-value crop, maintaining optimal nut quality is both an economic and agronomic priority (Memmi et al., 2016). Among the many factors influencing quality, irrigation stands out as a critical management tool, as it directly affects nut size, composition, and sensory attributes (Allakonon et al., 2022).

While irrigation is essential for sustaining pistachio production in areas with limited rainfall, its primary impact lies in how it shapes nut quality traits such as kernel development, lipid content, and nutrient composition (Kanber et al., 1993; Sedaghati & Hokmabadi, 2015). Both over and under-irrigation can compromise these quality parameters, leading to suboptimal texture, and nutritional value. Thus, the challenge is not merely to provide sufficient water, but to fine-tune irrigation regimes that maximize quality while minimizing resource use (Fader et al., 2016).

Climate change adds complexity by altering water availability and increasing the unpredictability of growing conditions (Rockström & Barron, 2007; Seneviratne et al., 2010; Wang & Zhang, 2011). As a result, the relationship between irrigation and pistachio quality is dynamic, requiring ongoing adaptation of management practices to sustain high-quality yields under shifting environmental pressures.

Traditional methods for assessing pistachio quality, such as chemical

* Corresponding author.
E-mail address: alvmarsa@itacyl.es (S. Álvarez).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.foodcont.2026.112011
Received 18 October 2025; Received in revised form 13 January 2026; Accepted 22 January 2026
Available online 23 January 2026
0956-7135/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

analyses, are accurate but time-consuming and destructive. In this context, precision agriculture technologies offer new opportunities for real-time, non-invasive monitoring (Anjum et al., 2023), such as unmanned aerial vehicles (UAVs) and remote sensing, (Alvarez-Vanhard et al., 2021; Olson & Anderson, 2021).

However, to fully capture the nuances of pistachio quality, advanced techniques like hyperspectral imaging (HSI) are increasingly important. HSI enables rapid, detailed assessment of nut composition and quality attributes, supporting both research and on-farm decision-making. Besides, HSI combined with Machine Learning (ML) offers a non-invasive, high-throughput solution for assessing pistachio quality attributes (Bonifazi et al., 2021; Mohammadi-Moghaddam et al., 2018). By capturing a wide range of wavelengths, HSI enables detailed characterization of biochemical components in pistachios, including proteins, carbohydrates, minerals, and fatty acids. ML algorithms such as Partial Least Squares (PLS), Support Vector Machines (SVM), and Extra Trees Regressor (ETR) can then leverage these spectral signatures to predict key quality parameters rapidly and accurately in food products (Saha & Manickavasagan, 2021).

Despite the proven potential of HSI and ML in quality prediction, limited attention has been given to pistachio quality prediction under different irrigation regimes. Recent studies have demonstrated the effectiveness of hyperspectral imaging (HSI) and machine learning (ML) for classifying pistachios according to origin and irrigation regimes, offering a rapid, non-destructive traceability method, as done by Martínez-Peña et al. (2024). However, that work focused primarily on classifying treatments and locations without addressing the quantitative prediction of specific nutritional parameters (e.g., minerals and fatty acids).

This study takes a step further by applying regression-based HSI and ML methods to precisely predict important pistachio components: nitrogen, fatty acids (oleic, linoleic, etc.), and other critical components. This approach aims to directly link irrigation and location to chemical composition changes in pistachios while also providing deeper insight into their impact on quality and overall yield performance. In this context, a portable VIS–NIR hyperspectral system operating in the 400–1000 nm range is employed, offering a relatively low-cost, field-deployable and user-friendly solution that is suitable for potential on-farm applications. Therefore, by integrating HSI with ML for pistachio quality assessment, this research offers a novel approach to assess irrigation practices and enhance orchard sustainability with the aim of: (1) assessing the combined influence of irrigation treatment and location on pistachio nutrient and fatty acid profiles, (2) developing and comparing predictive ML models using hyperspectral data, and (3) evaluating HSI as a rapid, non-destructive method for monitoring orchard performance and directing irrigation strategies. The findings could guide farmers and policymakers in making data-driven decisions that balance resource conservation with high-quality production, addressing growing concerns over water scarcity and climate variability.

2. Materials and methods

2.1. Plant material and image acquisition

Experiments were conducted in 2022 at two pistachio orchards located in Valladolid, Castilla y León, Spain. These orchards were situated in "Moraleja de las Panaderas" (M) 41.27048°, −4.81709°; WGS84) and "La Seca" (S) (41.44337°, −4.89781°; WGS84). Seven-year-old Moraleja and fifteen-year-old La Seca pistachio plants from the *Pistacia vera* cv. Kerman variety made up the plant material chosen for the research. These plants were grafted onto the *P. atlantica* × *P. integerrima* hybrid UCB rootstock. In order to maximize solar interception and ensure effective resource utilization, the orchard design used a 7 × 6 m triangle planting pattern oriented NE–SW. Cv. Peter was the male cultivar utilized in these orchards. The pistachio trees were subjected to two distinct irrigation treatments during their growth phase. The high irrigation treatment provided 50 % more water than the control treatment. The trees at "La Seca" received total irrigation volumes of 2750 m^3 ha^{-1} for the control treatment and 4660 m^3 ha^{-1} for the high irrigation treatment. In "Moraleja de las Panaderas", the control treatment received 844 m^3 per hectare, while the high treatment was given 1161 m^3 per hectare. This variation in water allocation was implemented to examine the effects of differential irrigation on the pistachio quality. In order to protect yield potential, standard agricultural procedures were adhered to, including the application of particular agrochemicals to control weeds, pests, and diseases.

Twenty trees, equally distributed across the two irrigation treatments and sites (five replicates per condition), were harvested when they reached maturity in October 2022. Post-harvest samples that were peeled, dried for 24 h at 60 °C to reduce mycotoxin concerns, and weighed were used to assess the yield (Yield, kg $tree^{-1}$).

The pistachios in each picture matched a particular sample that was gathered prior to harvest, which matched three bunches of pistachios from the 20 trees that were chosen in the various locations based on orientation (North, South, East, and West) and height (High and Low). However, some samples could not be taken from the upper or lower part of the trees at the Moraleja de las Panaderas area because of their age. Following the samples' processing, peeling, and 24 h drying at 60 °C, 158 pictures total, equivalent to 2818 pistachios, were captured. A total of 77 photographs were obtained from Moraleja de las Panaderas (37 from high irrigation treatment and 40 from control irrigation), and 80 from La Seca (40 from high irrigation treatment and 40 from control irrigation).

Images of the pistachios were acquired at a room with controlled temperature (23 °C) and lighting by using two halogen light sources, as they produce a continuous spectrum of light, from deep infrared to violet, covering the range of the SPECIM IQ hyperspectral camera (SPECIM IQ camera, Specim Spectral Imaging Oy Ltd, Oulu, Finland). The camera is equipped with a VNIR CMOS sensor, offering a 400–1000 nm spectral range and 204 spectral bands. It features a viewfinder camera with a 5 megapixel resolution, rescaled to 1280*960 pixels for image previews. On every image, a white reference was captured.

2.2. Image processing and data extraction

Python 3.10, along with Pandas and NumPy were employed for data manipulation, while Scikit-learn facilitated model development. Image processing was carried out using Scikit-image and NumPy. Finally, Seaborn and Matplotlib libraries were utilized for data visualisation. The reflectance of the images was corrected using white (Spectralon, NH, USA) with a 99 % reflectance and dark references, following Equation 1 (Sun, 2010), where 〖Im〗_Dark and 〖Im〗_White are the pixelwise average of the white and dark references, and 〖Im〗_RAW is the original image. Each image was corrected with its own white reference.

Equation 1. Hyperspectral image correction.

$$Im_{Refl} = \frac{Im_{RAW} - Im_{Dark}}{Im_{White} - Im_{Dark}}$$

The number of pistachios per hyperspectral image varied according to tree, treatment, and orchard location. Image analysis followed a standardized workflow, summarised in Fig. 1. First, principal component analysis (PCA; scikit-learn) was applied to each hyperspectral cube, and the first principal component image was used to enhance contrast between pistachios and background. This high-contrast image facilitated automatic segmentation by Otsu's global thresholding (Otsu, 1979) using scikit-image combined with Canny edge detection (OpenCV) to delineate object boundaries. The coordinates obtained from the edge map were then used to crop individual pistachio kernels from the original hyperspectral images, yielding one hyperspectral sub-cube per nut.

For each cropped pistachio, the mean spectrum was calculated over all pixels belonging to that nut using NumPy, thereby generating an

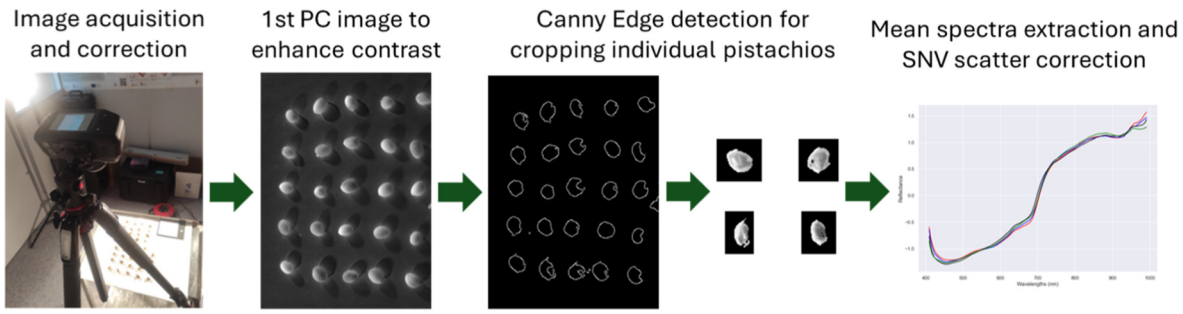

Fig. 1. Workflow of the image processing and spectral extraction.

averaged spectral dataset at the sample level. Using mean spectra is recommended in hyperspectral chemometrics because it reduces pixel-level noise, decreases the influence of local surface irregularities, and provides a more representative spectrum for each object, which in turn improves model robustness and interpretability. Several studies have shown that averaging spectra over regions of interest yields more stable calibration models and mitigates artefacts caused by heterogeneous illumination and surface geometry in fruit and nut applications. Importantly, averaging across the entire nut also alleviates the systematic reflectance differences that arise along the pistachio surface due to curvature and varying incidence angles of light, a well-known source of multiplicative scatter in imaging spectroscopy. By integrating spectra over the whole kernel, high-reflectance and low-reflectance pixels are balanced, so that each sample is characterised by a single spectrum that reflects its overall composition rather than local shading effects (Amigo et al., 2015; H. Wang et al., 2015).

Prior to modelling, the averaged spectra was pre-processed to minimize remaining physical scatter effects. Standard normal variate (SNV) was applied (Equation 2) to each spectrum, which rescales it to zero mean and unit variance by subtracting its own mean and dividing by its standard deviation. This transformation is widely used in NIR and hyperspectral studies to reduce multiplicative scatter and path-length variability between samples, thereby allowing models to emphasize genuine chemical differences instead of measurement geometry or particle-size effects (Fearn et al., 2009).

Equation 2. Standard Normal Variate sctter correction.

$$Corrected\ spectra = \frac{Original\ spectra - Mean\ spectra}{Standard\ deviation}$$

2.3. Quality parameters obtention

The physicochemical characteristics of a sample of pistachio nuts taken from the trees under study were evaluated to determine their nutritional value, each sample analysed corresponded to the pistachios captured in each image. On the one hand, the fat content (Fat, g 100 g^{-1}) was extracted using the Ankom sealed bag system and petroleum ether, while the fiber (g 100 g^{-1}) was determined using the enzymatic–-gravimetric method by sequential enzymatic digestion with α-amylase, protease, and amyloglucosidase. Conversely, the protein content (PROT, g 100 g^{-1}) was computed using the Dumas combustion method for nitrogen quantification and multiplied by the value obtained with the coefficient 6.25 (Nielsen, 2010). The carbohydrate content (CBH, g 100 g_{-1}) was calculated using the difference of other major components (CBH = 100 − Ash-Humidity-PROT-Fat), where the ash and humidity content were measured using gravimetric methods.

Carbon (C, g kg^{-1}), calcium (Ca, mg kg^{-1}), magnesium (Mg, mg kg^{-1}), phosphorus (P, mg kg^{-1}), sodium (Na, mg kg^{-1}), iron (Fe, mg kg^{-1}), and zinc (Zn, mg kg^{-1}) were also measured. Therefore, 500 mg of dry material was digested at 200 °C in a Teflon container using the ETHOUS UP microwave digestion system (Milestone, Sorisole, Italy). After cooling to room temperature, the solutions were diluted. Finally, an Agilent Technologies Varian 720-ES inductively coupled plasma-optical emission spectrometer (ICP-OES; Santa Clara, CA, USA) was used to quantify the nutrient content. Following dry acid digestion, gas chromatography with a flame ionization detector (GC-FID) was used to quantify the proportion of various fatty acids, including palmitic, palmitoyl, stearic, oleic and linolenic.

2.4. Statistical analysis

Statgraphics Centurion 19 (Statistical Graphics Corp., Rockville, MA, USA) was used to perform a multifactor analysis of variance (ANOVA) and Tukey's multiple range test to the instrumental pistachio quality data. A Pearson correlation matrix was also performed using Python and the library Seaborn in order to study the linear relation between the measured pistachio components. Statistically significant differences were considered when $p < 0.05$.

2.5. Model creation

ML models were trained using the mean spectra extracted from each pistachio to predict several key quality parameters: Ca, P, Fe, Mg, N, Na, Zn, C, palmitic, palmitoyl, stearic, oleic linoleic, linolenic, fat, ash, CBH, humidity, protein and fibre content. Each image provided only one predictive value, meaning all pistachios within the same image shared identical quality parameters.

Three widely used regression algorithms were implemented in Python using the scikit-learn library: Partial Least Squares Regression (PLS-R), Support Vector Regression (SVR, referred to here as SVM), and Extra Trees Regressor (ETR). PLS-R is a latent-variable method that projects the high-dimensional spectral matrix onto a small number of orthogonal components that maximize covariance with the response, making it particularly suitable for collinear hyperspectral data and routinely used in fruit-quality spectroscopy. SVR maps the input spectra into a high-dimensional feature space via a kernel function and fits a sparse regression function that controls model complexity through margin and regularization parameters, widely used predicting physicochemical attributes from spectral data in horticultural applications. ETR is an ensemble of randomized decision trees that aggregates many de-correlated trees to capture non-linear relationships while remaining relatively interpretable and robust to noise (Mü; ller & Guido, 2016; Wieme et al., 2022).

The dataset comprised 2818 mean VIS–NIR spectra, each corresponding to a single pistachio nut. Given this moderate sample size and the primary objective of building accurate yet interpretable predictive models for compositional quality traits, established chemometric and machine-learning methods (PLS-R, SVR, tree ensembles) were preferred over deep neural networks such as multilayer perceptrons, one-dimensional convolutional networks or transformer architectures. Deep learning models typically require substantially larger labelled datasets and careful regularization to avoid overfitting in hyperspectral applications, and, for small-to-medium spectral datasets, conventional models like PLS-R or kernel-based regressors often match or exceed the

performance of more complex neural networks while being easier to tune and interpret. For these reasons, and to ensure reproducibility and transparency in an applied pistachio-quality context, the modelling strategy focused on these three established algorithms (Chollet, 2021; Hasanzadeh et al., 2022).

Before model training, the full dataset was randomly partitioned into a training set containing 70 % of the spectra (1972 samples) and an independent validation set comprising the remaining 30 % (846 samples), using scikit-learn's 'train_test_split' function. A random split was selected because the two orchard locations were managed under similar conditions and were not intended to represent independent production environments; in this context, pooling all samples and performing a random division provides an unbiased estimate of model performance on new pistachios from the same target population and is widely adopted in fruit-quality spectroscopy studies with comparable sample sizes. Location-based splitting can be advantageous when sites represent clearly distinct external domains, but with only two orchards and limited per-site sample numbers it would substantially reduce the effective training set and risk unstable estimates of generalisation, whereas the chosen strategy balances training data availability with an independent assessment set (Beleites & Salzer, 2008; Ramezan et al., 2019). Future research should, however, include additional years and a larger number of orchards so that fully independent external datasets (e.g. based on different locations or production seasons) can be reserved for validation, thereby enabling a more rigorous assessment of the spatial and temporal robustness of the proposed models.

Model hyperparameters were optimised using scikit-learn's 'GridSearchCV'. In this procedure, a predefined grid of candidate hyperparameter combinations is systematically evaluated using only the training data. Here, 10-fold cross-validation was applied within the training set: the training data was partitioned into 10 folds, models were iteratively fitted on 9 folds and validated on the remaining fold, and performance was averaged across folds for each parameter combination. The configuration yielding the lowest cross-validated root mean squared error (RMSE) and highest coefficient of determination (R^2) was selected as optimal for each algorithm. Table 1 shows the selected hyperparameters of the different models. Finally, new models were refitted on the entire training set using these best hyperparameters and their predictive performance was evaluated once on the independent validation set.

The trained models were evaluated using R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). R^2 indicates the proportion of variance in the dependent variable explained by the independent variables, MAE measures the average absolute difference between predicted and actual values, and RMSE provides a measure of the average absolute error between predicted and actual values, expressed in the same units as the original data.

3. Results

3.1. Quality results

3.1.1. Nutrients

Significant differences ($p < 0.001$) in the concentrations of Ca, P, Fe, Mg, N, Na, Zn, and C were found in pistachio samples across different locations, treatments, and their combinations (Table 2). The location appeared to have a pronounced effect, with samples from Moraleja de las Panaderas generally having higher concentrations of most nutrients than those from La Seca. This pattern was particularly evident for N and P (3.9675 vs 3.6717 g kg^{-1} for N, and 0.5886 vs 0.5153 mg kg^{-1} for P), suggesting that environmental conditions at Moraleja de las Panaderas may favour the accumulation of these elements. However, Na levels were slightly higher in La Seca samples (17.6496 vs 16.9575 mg kg^{-1}), indicating a location-specific variation in nutrient uptake or soil composition.

The influence of treatment was also present, with high-treatment samples showing higher nutrient concentrations compared to control-treatment samples. This trend was found in Fe, Na or Zn (35.6716 vs 34.1064 mg kg^{-1} for Fe, 18.5694 vs 16.0378 mg kg^{-1} for Na, and 19.8796 vs 17.4038 mg kg^{-1} for Zn), suggesting that intensive management practices enhance the uptake or accumulation of these nutrients. The interaction between location and treatment highlights the complexity of these effects. For instance, Moraleja de las Panaderas high-treatment samples exhibited the highest Ca, P, and N levels (0.1154 mg kg^{-1} for Ca, and 0.6430 mg kg^{-1} for P), while La Seca control-treatment samples generally showed lower concentrations. These findings suggest that both environmental factors and agronomic practices interact to shape the nutrient profile of pistachios.

However, carbon content showed less pronounced variations than other nutrients but still displays significant differences across treatments and locations. This suggests that carbon levels are influenced by broader metabolic or environmental factors rather than specific agronomic interventions.

Fig. 2 provides a graphical summary of the mineral composition described numerically in Table 2. The boxplots for Ca, P, Fe, Mg, N, Na, Zn, and C suggest limited variability in mineral concentrations among samples. Although minor differences in spread are evident, particularly for Fe and Na, the overall pattern denotes a stable and homogeneous mineral profile.

3.1.2. Fat

The results of the ANOVA and Tukey tests shown in Table 3 reveal significant differences ($p < 0.001$) in the fatty acid composition and overall fat content of pistachios across locations, treatments, and their combinations. Oleic acid content was higher in pistachios from Moraleja de las Panaderas compared to La Seca (58.495 vs 53.588 % of Fat), indicating that environmental conditions in Moraleja de las Panaderas may favour the accumulation of this monounsaturated fatty acid. On the other hand, linoleic acid levels were higher in La Seca samples, suggesting a location-dependent variation in the balance between unsaturated fatty acids (29.904 vs 25.012 % of Fat). Though present in lower concentrations, Linolenic acid also showed significant location differences (0.609 in La Seca vs 0.535 % of Fat in Moraleja de las Panaderas).

Water treatment effects were observed, with high-treatment samples generally showing increased levels of palmitic and palmitoyl acids compared to control-treatment samples (13.172 vs 12.449 % of Fat for palmitic and 1.558 vs 1.445 % of Fat for palmitoyl). This suggests that agronomic practices influence the accumulation of saturated fatty acids. However, oleic acid content was slightly lower in high-treatment samples (56.426 vs 55.656 % of Fat). However, no statistical differences were found regarding linoleic content in water treatment. The interaction between location and treatment further emphasizes the complexity of these effects. Moraleja de las Panaderas control-treatment samples exhibited the highest oleic acid levels (58.732 % of Fat), while La Seca high-treatment samples showed elevated linoleic acid (0.600 % of Fat) and overall fat content (48.370 g 100 g). These patterns suggest that both environmental conditions and management practices also interact to influence the fatty acid composition and total fat content of pistachios.

The numerical data reported in Table 3 are more intuitively represented in Fig. 3, which shows the boxplots for total fat content and the major fatty acids. It shows narrow interquartile ranges, reflecting a consistent lipid profile among samples. Oleic and linoleic acids display the highest median proportions and slightly wider variability, whereas palmitic, palmitoyl, stearic, and linolenic acids present more compact distributions around their central values.

3.1.3. Others

Table 4 shows the results of the ANOVA and Tukey tests on ash, CBH, humidity, protein, and fiber content of pistachios across locations, treatments, and their combinations. The location appeared to substantially influence these parameters ($p < 0.001$), with Moraleja de las

Table 1
Hyperparameters for the ML models used for each prediction.

Regression models hyperparameters						
Models	PLS		ETR		SVM	
Ash	Components:	27	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	10
			Max_depth:	24	Epsilon:	0.001
C	Components:	21	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	100
			Max_depth:	1	Epsilon:	0.001
Ca	Components:	22	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	0.1
			Max_depth:	1	Epsilon:	0.001
CBH	Components:	25	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
P	Components:	22	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
Fat	Components:	21	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	10
			Max_depth:	1	Epsilon:	0.001
Fe	Components:	18	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
Fiber	Components:	18	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	100
			Max_depth:	1	Epsilon:	0.001
Humidity	Components:	33	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	10
			Max_depth:	24	Epsilon:	0.001
Linoleic	Components:	25	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
Linolenic	Components:	21	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	0.1
			Max_depth:	24	Epsilon:	0.001
Mg	Components:	20	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	0.1
			Max_depth:	24	Epsilon:	0.001
N	Components:	23	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	10
			Max_depth:	24	Epsilon:	0.001
Na	Components:	17	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	100
			Max_depth:	1	Epsilon:	0.001
Oleic	Components:	25	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
Palmitic	Components:	30	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	10
			Max_depth:	1	Epsilon:	0.001

(*continued on next page*)

Table 1 (*continued*)

Regression models hyperparameters						
Models	PLS		ETR		SVM	
Palmitoyl	Components:	32	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	10
			Max_depth:	24	Epsilon:	0.001
Protein	Components:	23	N_estimators:	638	Kernel:	Poly
			Min_samples_split:	5	Gamma:	Scale
			Min_samples_leaf:	4	C:	100
			Max_depth:	24	Epsilon:	0.001
Stearic	Components:	19	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	1
			Max_depth:	1	Epsilon:	0.001
Zn	Components:	21	N_estimators:	1685	Kernel:	Poly
			Min_samples_split:	20	Gamma:	Scale
			Min_samples_leaf:	1	C:	500
			Max_depth:	1	Epsilon:	0.001

Panaderas samples generally showing higher ash and protein content than La Seca samples (3.4168 vs 3.0980 g 100 g^{-1} for ash and 24.7977 vs 22.9496 g 100 g^{-1} for protein content). On the other hand, La Seca samples showed higher CBH levels (22.0347 vs 19.8530 g 100 g^{-1}), suggesting that environmental conditions in La Seca may favour carbohydrate accumulation. Humidity content was also higher in Moraleja de las Panaderas samples (4.6724 vs 4.2991 g 100 g^{-1}), indicating potential differences in water retention or environmental moisture between the two locations. Fiber content showed no statistical differences between locations.

Water treatment effects were evident for most parameters ($p < 0.001$), with high-treatment samples generally exhibiting increased ash and fiber content compared to control-treatment samples. However, in the case of CBH, no statistical differences were found regarding water treatment. The interaction between location and treatment highlights the complexity of these effects, and both environmental conditions and agronomic practices interact to influence the biochemical composition of pistachios. For example, Moraleja de las Panaderas high-treatment samples showed the highest ash and protein levels (3.6176 g 100 g^{-1} for ash, and 25.6645 g 100 g^{-1} for protein), while La Seca control-treatment samples exhibited elevated CBH content (22.6875 g 100 g^{-1}).

Fiber content showed variations across treatments and their interactions with location, with high-treatment samples generally showing higher values. However, fiber was not influenced by location, and no statistical differences were found. These results underscore the importance of environmental factors and management practices in shaping the nutritional profile of pistachios.

The numerical results presented in Table 4 are more clearly illustrated in Fig. 4, which displays the boxplots for ash, CBH, humidity, protein, and fiber contents. Fig. 4 shows relatively narrow interquartile ranges for all variables, indicating a generally homogeneous chemical composition among samples. Protein and CBH exhibit slightly higher variability than the other components, whereas ash, fiber, and humidity display more compact distributions around their medians.

3.2. Parameter correlations

3.2.1. Nutrients

The correlation matrix (Fig. 5) revealed strong relationships between the factors influencing pistachio yield and nutrient composition. As expected, yield was positively associated with location and water treatment (0.6 and 0.65, respectively), suggesting that both environmental and water treatment play significant roles in productivity. As expected, N content appeared to be a critical factor among the nutrients, showing a strong positive correlation with yield (0.71). Inversely, Fe and Zn contributed negatively to productivity (−0.70 and −0.75). P highly correlated positively with Mg (0.92). Zn and Fe were also strongly correlated (0.77), and Ca with P and Mg (0.73 and 0.71, respectively). C was negatively associated with several nutrients, including Ca and Mg (−0.7 and −0.62). These results further highlight the complexity of combined environmental and management influences on pistachio quality.

3.2.2. Fat

The correlation matrix shown in Fig. 6 revealed significant relationships between location, treatment, their interaction, yield, and fat composition of pistachio samples. Yield exhibited positive correlations among the fat components with stearic acid and oleic acids (0.61) while showing negative associations with linoleic, palmitoyl and palmitic (−0.64, −0.81, −0.79, respectively). This suggests that specific shifts in the fatty acid profile of pistachios accompany higher yields. Besides, the interplay between fatty acids highlights notable patterns. Oleic acid was strongly negatively correlated with linoleic and linolenic acids (−0.98 and −0.77), reflecting an inverse relationship between these two major components of pistachio oil. Linoleic and linolenic were also correlated (0.7), and stearic was inversely correlated with linoleic and linoleic acids (−0.67 and −0.63) and positively correlated with oleic (0.73). On the other hand, overall fat content showed weak or negligible correlations with most parameters, indicating that total fat levels were less influenced by location, treatment, or yield compared to specific fatty acid profiles. The results underscore the complex interactions between environmental factors, management practices, and biochemical traits in pistachios.

3.2.3. Other

The correlation matrix in Fig. 7 highlights the relationships between location, treatment, their interaction, yield, and the analysed parameters of ash, CBH, humidity, protein, and fiber in pistachio samples. Location exhibited positive correlations with humidity, protein and yield (0.83, 0.75 and 0.6, respectively) and negative correlations with CBH (−0.84). Besides, as expected, the water treatment positively correlated with humidity (0.6). Yield showed positive relations with humidity and protein (0.67 and 0.71) and negative with CBH (−0.68). CBH was inversely correlated with the humidity and protein content (−0.63 and −0.82). As previously commented, humidity content was positively correlated with protein (0.76) and negatively correlated with CBH, indicating that higher moisture levels may dilute certain biochemical components. Protein content demonstrated strong positive correlations with yield, CBH and humidity, suggesting that higher protein levels were associated with higher water content and increased carbohydrate content. Fiber showed weak or no correlations with most parameters, indicating limited influence from environmental or

Table 2
ANOVA and Tukey test of the nutrient measurements of the pistachios.

	Ca (mg kg^{-1})	P (mg kg^{-1})	Fe (mg kg^{-1})	Mg (mg kg^{-1})	N (g kg^{-1})	Na (mg kg^{-1})	Zn (mg kg^{-1})	C (g kg^{-1})
ANOVA test	***	***	***	***	***	***	NS	***
S	0.1006 ± 0.0002 ^{b}	0.5153 ± 0.0008 ^{b}	36.6136 ± 0.0560 ^{a}	0.1065 ± 0.0001 ^{b}	3.6717 ± 0.0035 ^{b}	17.6496 ± 0.0877 ^{a}	18.6146 ± 0.0306	60.9214 ± 0.0081 ^{a}
M	0.1009 ± 0.0002 ^{a}	0.5886 ± 0.0008 ^{a}	33.1644 ± 0.0620 ^{b}	0.1197 ± 0.0001 ^{a}	3.9675 ± 0.0038 ^{a}	16.9575 ± 0.0971 ^{b}	18.6689 ± 0.0339	60.5064 ± 0.0090 ^{b}
ANOVA test	***	***	***	***	***	***	***	***
H	0.1070 ± 0.0002 ^{a}	0.5744 ± 0.0009 ^{a}	35.6716 ± 0.6610 a	0.1147 ± 0.0002 ^{a}	3.7562 ± 0.0041 ^{b}	18.5694 ± 0.1036 ^{a}	19.8796 ± 0.0361 ^{a}	60.8430 ± 0.0100 ^{a}
C	0.1027 ± 0.0001 ^{b}	0.5295 ± 0.0007 ^{b}	34.1064 ± 0.0510 ^{b}	0.1115 ± 0.0001 ^{b}	3.8830 ± 0.0031 ^{a}	16.0378 ± 0.0798 ^{b}	17.4038 ± 0.0279 ^{b}	60.5848 ± 0.0073 ^{b}
ANOVA test	***	***	***	***	***	***	***	***
S H	0.0986 ± 0.0003 ^{c}	0.5057 ± 0.0011 ^{d}	35.7871 ± 0.0792 ^{b}	0.1046 ± 0.0002 ^{d}	3.6834 ± 0.0049 ^{c}	19.7533 ± 0.1240 ^{a}	17.7689 ± 0.0433 ^{c}	61.3910 ± 0.0114 ^{a}
S C	0.1027 ± 0.0003 ^{b}	0.5249 ± 0.0011 ^{c}	37.4401 ± 0.0791 ^{a}	0.1084 ± 0.0002 ^{c}	3.6600 ± 0.0049 ^{d}	15.5459 ± 0.1239 ^{d}	21.9904 ± 0.0580 ^{a}	60.4519 ± 0.0114 ^{c}
M H	0.1154 ± 0.0003 ^{a}	0.6430 ± 0.0014 ^{a}	35.5561 ± 0.1060 ^{b}	0.1248 ± 0.0003 ^{a}	3.8290 ± 0.0066 ^{b}	17.3855 ± 0.1661 ^{b}	19.4603 ± 0.0433 ^{b}	60.2951 ± 0.0153 ^{d}
M C	0.1027 ± 0.0002 ^{b}	0.5341 ± 0.0009 ^{b}	30.7727 ± 0.0642 ^{c}	0.1146 ± 0.0002 ^{b}	4.1062 ± 0.0040 ^{a}	16.5296 ± 0.1006 ^{c}	15.3474 ± 0.0351 ^{d}	60.7177 ± 0.0092 ^{b}

S, La Seca; M, Moraleja de las Panaderas; H, High; C, Control; NS, not significant at $p < 0.05$.
Values followed by the same letter within the same column were not significantly different ($p < 0.05$), according to Tukey's least significant difference test.
*** Significant at $p < 0.001$.

agronomic factors in this dataset, highlighting the intricate relationship between environmental factors, farming practices, and biochemical characteristics in pistachios.

3.3. Quality predictions

The SNV corrected VIS–NIR spectra (Fig. 8) showed the greatest variability in two main regions: the visible red edge between approximately 650 and 750 nm and the short wave NIR region from about 800 to 1000 nm. In the red edge zone, the spectra exhibit a steep change in reflectance and a clear spread between samples, which is consistent with the influence of chlorophyll and carotenoid absorptions and their known sensitivity to nitrogen status, protein content and tissue maturity in plant materials (Peng et al., 2014; Zhang et al., 2016). In the 800–1000 nm interval, the curves diverge again, particularly around 900–980 nm, where overtones and combination bands of O–H, C–H and C–O groups occur; this region is widely exploited to quantify moisture, total fat and organic carbon in agricultural products and soils (Ojha et al., 2024; Peng et al., 2014; Wu & Sun, 2013).

These spectral characteristics provide a chemical rationale for the prediction of the measured parameters in pumpkins. Minerals such as Ca, P, Fe, Mg, Na and Zn, as well as ash and total C, do not show strong NIR bands themselves but are closely associated with organic matrices (proteins, cell walls, storage tissues) and water, which govern absorption and scattering in the 650–1000 nm range; this indirect relationship has been successfully used to model macro and micronutrients from VIS–NIR data in plant tissues and soils (Khadivi et al., 2025b; Y. Liu et al., 2019). Total fat and the main fatty acids (oleic, linoleic, linolenic, palmitic, stearic and palmitoyl containing lipids) predominantly affect C–H overtones that are strongest between about 850 and 1000 nm, and previous work on vegetable oils and nuts has linked features near 900–930 nm and 960–980 nm with variations in fatty acid composition, supporting the use of this region for predicting lipid related traits in pumpkin flesh or seeds (Jensen et al., 2001; Teixeira & Sousa, 2019). Water content (humidity) strongly influences the band near 970 nm, corresponding to the first overtone of O–H stretching, which is a standard predictor of moisture in grains, seeds and pistachios (Mohammadi-Moghaddam et al., 2017), whereas protein is commonly modeled from N–H and C–H overtones distributed across 700–1000 nm, including the red edge and short wave NIR regions where the largest variation is observed (Kimiya et al., 2013; Wojciechowski et al., 2016). Finally, carbohydrates and fiber, dominated by polysaccharidic C–H and O–H bonds, contribute substantially to absorptions between 850 and 1000 nm, a range routinely used to quantify starch and structural carbohydrates in seeds and flours, which explains the sensitivity of this portion of the spectrum to CBH and fiber content (Chen et al., 2013; Paz-Kagan et al., 2020; Zhou et al., 2025).

Taken together, the clear spectral differentiation in these chemically meaningful regions indicates a strong relationship between VIS–NIR signatures and the studied quality attributes, and therefore supports the development of chemometric predictive models based on these spectra to estimate the physicochemical parameters of pumpkins.

3.3.1. Nutrient predictions

Table 5 shows the results of the predictive models (PLS, SVM, and ETR). Across all nutrients, PLS generally exhibited higher R^2 values and lower error metrics (MAE and RMSE) compared to SVM and ETR, suggesting its superior performance in capturing the relationships between predictors and nutrient outcomes. SVM showed slightly lower R^2 values and marginally higher errors than PLS. These results are aligned with the correlations previously studied. For instance, nutrients such as N and Zn, which exhibited strong positive correlations with yield, were predicted with higher accuracy across all models (R^2 of 0.75 and 0.81 with the PLS for N and Zn, respectively). In contrast, nutrients with weaker or more complex relationships, such as Na or Mg, showed lower predictive performance, especially in the case of Na, which showed the lowest

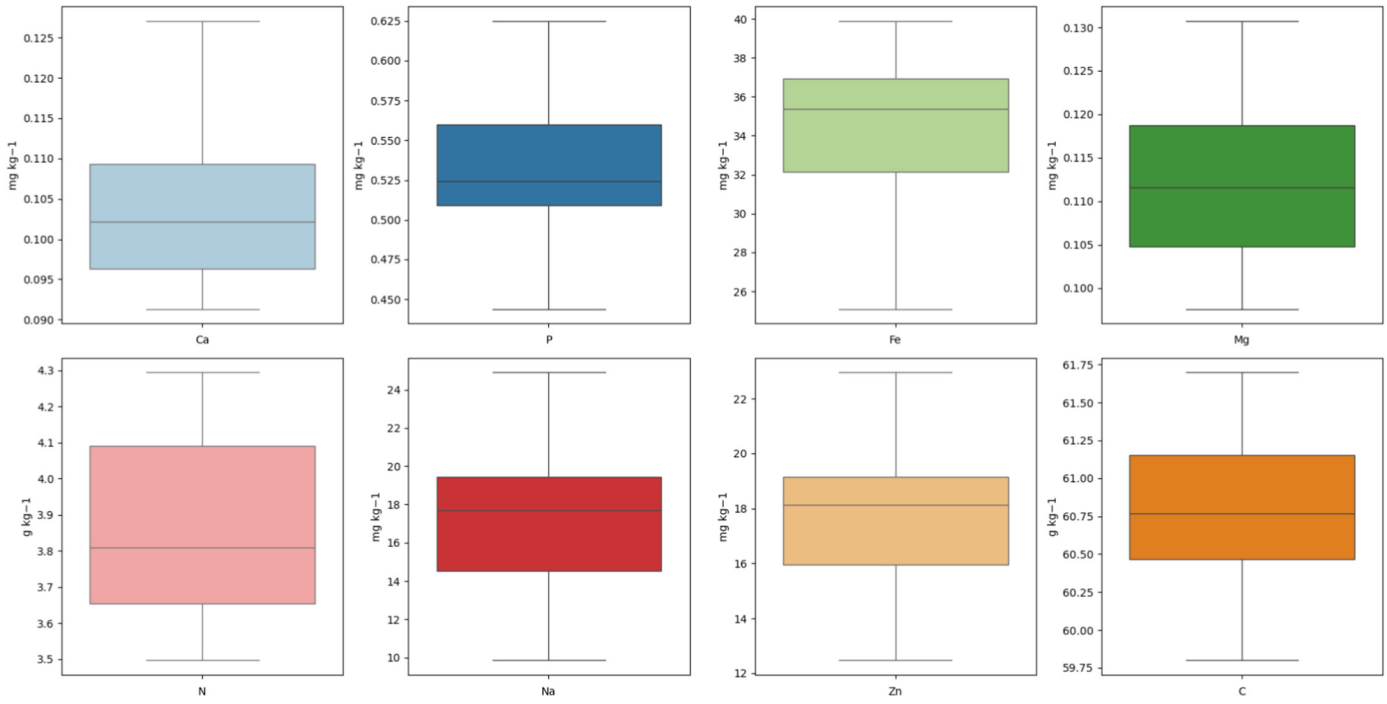

Fig. 2. Boxplots of the different pistachio nutrients analysed.

Table 3
ANOVA and Tukey test of the fat measurements of the pistachios.

	Fat (g 100 g)	Oleic (% of Fat)	Linolenic (% of Fat)	Linoleic (% of Fat)	Palmitic (% of Fat)	Palmitoyl (% of Fat)	Steraic (% of Fat)
ANOVA test	***	***	***	***	***	***	***
S	47.619 ± 0.022 ^{a}	53.588 ± 0.021 ^{b}	0.609 ± 0.001 ^{a}	29.904 ± 0.024 ^{a}	12.795 ± 0.006 ^{b}	1.568 ± 0.002 ^{a}	1.025 ± 0.002 ^{b}
M	47.260 ± 0.025 ^{b}	58.495 ± 0.023 ^{a}	0.535 ± 0.001 ^{b}	25.012 ± 0.027 ^{b}	12.825 ± 0.006 ^{a}	1.434 ± 0.002 ^{b}	1.124 ± 0.002 ^{a}
ANOVA test	***	***	***	NS	***	***	***
H	47.860 ± 0.026 ^{a}	55.656 ± 0.025 ^{b}	0.577 ± 0.001 ^{a}	27.469 ± 0.028	13.172 ± 0.007 ^{a}	1.558 ± 0.002 ^{a}	1.043 ± 0.002 ^{b}
C	47.018 ± 0.020 ^{b}	56.426 ± 0.019 ^{a}	0.567 ± 0.001 ^{b}	27.448 ± 0.022	12.449 ± 0.005 ^{b}	1.445 ± 0.001 ^{b}	1.106 ± 0.002 ^{a}
ANOVA test	***	***	***	***	***	***	***
S H	48.370 ± 0.032 ^{a}	53.055 ± 0.030 ^{d}	0.600 ± 0.001 ^{b}	30.466 ± 0.034 ^{a}	12.841 ± 0.008 ^{b}	1.537 ± 0.003 ^{c}	1.018 ± 0.003 ^{d}
S C	46.867 ± 0.032 ^{d}	54.120 ± 0.030 ^{c}	0.618 ± 0.001 ^{a}	29.343 ± 0.034 ^{b}	12.749 ± 0.008 ^{c}	1.600 ± 0.003 ^{a}	1.031 ± 0.003 ^{c}
M H	47.351 ± 0.042 ^{b}	58.257 ± 0.040 ^{b}	0.554 ± 0.001 ^{c}	24.471 ± 0.046 ^{d}	13.501 ± 0.011 ^{a}	1.578 ± 0.003 ^{b}	1.068 ± 0.004 ^{b}
M C	47.169 ± 0.026 ^{c}	58.732 ± 0.024 ^{a}	0.517 ± 0.001 ^{d}	25.553 ± 0.028 ^{c}	12.148 ± 0.007 ^{d}	1.290 ± 0.002 ^{d}	1.180 ± 0.002 ^{a}

S, La Seca; M, Moraleja de las Panaderas; H, High; C, Control; NS, not significant at $p < 0.05$.
Values followed by the same letter within the same column were not significantly different ($p < 0.05$), according to Tukey's least significant difference test.
*** Significant at $p < 0.001$.

performance, with R^2 of prediction of 0.20 for Na and 0.63 and Mg, respectively. P also showed a high predictive value with an R^2 of 0.70 with the PLS, whereas, in contrast, Ca, Fe and C showed R^2 scores of 0.44, 0.65 and 0.55, respectively.

3.3.2. *Fat predictions*

The results from the predictive models on the composition of pistachio samples are shown in Table 6. As in nutrient prediction, PLS demonstrated the highest accuracy across most parameters. These results are supported by its higher R^2 values and lower error metrics, indicating that PLS effectively captured the linear relationships between predictors and fatty acid concentrations. Particularly effective in predicting oleic acid, linoleic acid, palmitic and palmitoyl with R^2 values of 0.91, 0.87, 0.82 and 0.82, respectively. SVM performed similarly, achieving similar accuracies for oleic and linoleic acid predictions (R^2 of 0.90 and 0.86, respectively), though its performance was slightly behind PLS for other parameters. ETR, on the other hand, showed lower predictive power across all metrics except linolenic acid, with the highest R^2 of 0.75.

The model performances aligned with the observed relationships between variables. For instance, oleic acid and linoleic acid exhibited strong correlations with location and yield in the matrix, which likely contributed to their higher predictability across models. In contrast, overall fat content showed weaker correlations with other factors in the matrix, which may explain its lower predictive accuracy.

3.3.3. *Other predictions*

Table 7 shows the results of the predictive models on CBH, humidity, protein, and fiber content predictions in pistachio samples. PLS achieved the highest accuracy across most parameters, particularly in ash, CBH and humidity, with R^2 values of 0.81, 0.87 and 0.84, respectively. Protein content achieved an R^2 of 0.75, whereas fiber prediction was unsuccessful with any of the three models. SVM also performed similarly to PLS for ash, CBH, humidity and protein predictions but showed slightly reduced performance for other parameters.

The performance of the models reflected the strength of the

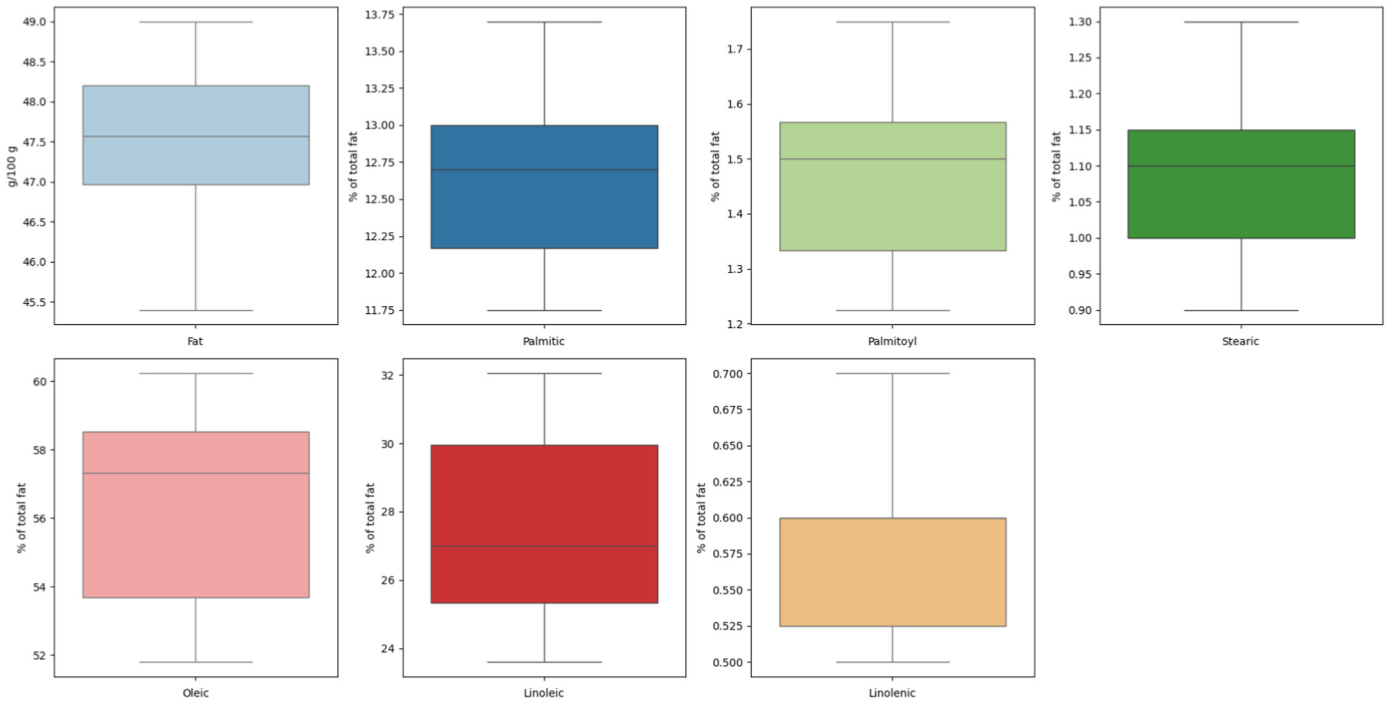

Fig. 3. Boxplots of the different pistachio fats analysed.

Table 4
ANOVA and Tukey test of ash, CBH, humidity, protein and fiber measurements of the pistachios.

	Ash (g 100 g^{-1})	CBH (g 100 g^{-1})	Humidity (g 100 g^{-1})	Protein (g 100 g^{-1})	Fiber (g 100 g^{-1})
ANOVA test	***	***	***	***	NS
S	3.0980 ± 0.0023 ^{b}	22.0347 ± 0.0135 ^{a}	4.2991 ± 0.0024 ^{b}	22.9496 ± 0.0216 ^{b}	11.9425 ± 0.0363
M	3.4168 ± 0.0025 ^{a}	19.8530 ± 0.0150 ^{b}	4.6724 ± 0.0026 ^{a}	24.7977 ± 0.0239 ^{a}	11.9936 ± 0.0402
ANOVA test	***	NS	***	***	***
H	3.3363 ± 0.0027 ^{a}	20.9486 ± 0.0160	4.3776 ± 0.0028 ^{b}	23.4771 ± 0.0255 ^{b}	12.5638 ± 0.0429 ^{a}
C	3.1786 ± 0.0021 ^{b}	20.9391 ± 0.0123	4.5940 ± 0.0021 ^{a}	24.2702 ± 0.0197 ^{a}	11.3723 ± 0.0330 ^{b}
ANOVA test	***	***	***	***	***
S H	3.0550 ± 0.0033 ^{d}	21.3819 ± 0.0191 ^{b}	4.1701 ± 0.0033 ^{d}	23.0232 ± 0.0306 ^{c}	12.3244 ± 0.0514 ^{b}
S C	3.1411 ± 0.0033 ^{c}	22.6875 ± 0.0191 ^{a}	4.4281 ± 0.0033 ^{c}	22.8760 ± 0.0306 ^{d}	11.5606 ± 0.0513 ^{c}
M H	3.6176 ± 0.0044 ^{a}	20.5153 ± 0.0256 ^{c}	4.5850 ± 0.0045 ^{b}	25.6645 ± 0.0248 ^{a}	12.8032 ± 0.0688 ^{a}
M C	3.2160 ± 0.0026 ^{b}	19.1908 ± 0.0155 ^{d}	4.7598 ± 0.0027 ^{a}	23.9310 ± 0.0409 ^{b}	11.1841 ± 0.0417 ^{d}

S, La Seca; M, Moraleja de las Panaderas; H, High; C, Control; NS, not significant at $p < 0.05$.
Values followed by the same letter within the same column were not significantly different ($p < 0.05$), according to Tukey's least significant difference test.
*** Significant at $p < 0.001$.

relationships observed between the variables. For example, CBH and protein content showed strong positive correlations with yield and other parameters in the dataset, which likely enhanced their predictability. In contrast, fiber exhibited weaker correlations with most variables in the dataset, which could explain its lower predictive accuracy.

4. Discussion

The results presented in this study highlight how environmental conditions (location) and agronomic practices (irrigation regimes) significantly shape the nutritional and fatty acid profiles of pistachios, as well as their overall quality. Similar to previous research on hyperspectral imaging (HSI) and Machine Learning (ML) applied to nuts and seeds (Gilman, 2021; Goldhamer & Beede, 2004), our findings confirm that both location and water management introduce distinct compositional traits in pistachios, which can be detected and quantified accurately using spectral data. This work expands upon an earlier classification-focused study on pistachios (Martínez-Peña et al., 2024), where the primary goal was to distinguish pistachio origin and irrigation treatments. In this study, the focus was expanded by quantifying key nutrients and fatty acids, providing deeper insight into how specific agronomic conditions drive compositional changes.

One of the most important outcomes was the marked influence of location, with Moraleja de las Panaderas samples consistently exhibiting higher N, P, oleic acid, protein, and ash contents. Environmental factors such as soil fertility, local climatic conditions, and orchard age may contribute to this enhanced nutrient accumulation (Li et al., 2024). By contrast, samples from La Seca showed moderately lower nutrient levels but slightly higher Na, possibly indicating differences in soil composition or water salinity. Notably, irrigation treatment also played a central role, with high-irrigation pistachios displaying elevated levels of Fe, Zn, and saturated fatty acids. This aligns with other orchard crop studies showing that water availability can enhance mineral uptake or alter metabolic pathways, ultimately modifying the fruit's biochemical composition (Kudoyarova et al., 2015; Rouphael et al., 2012).

Additionally, the correlation analysis provided deeper insight into

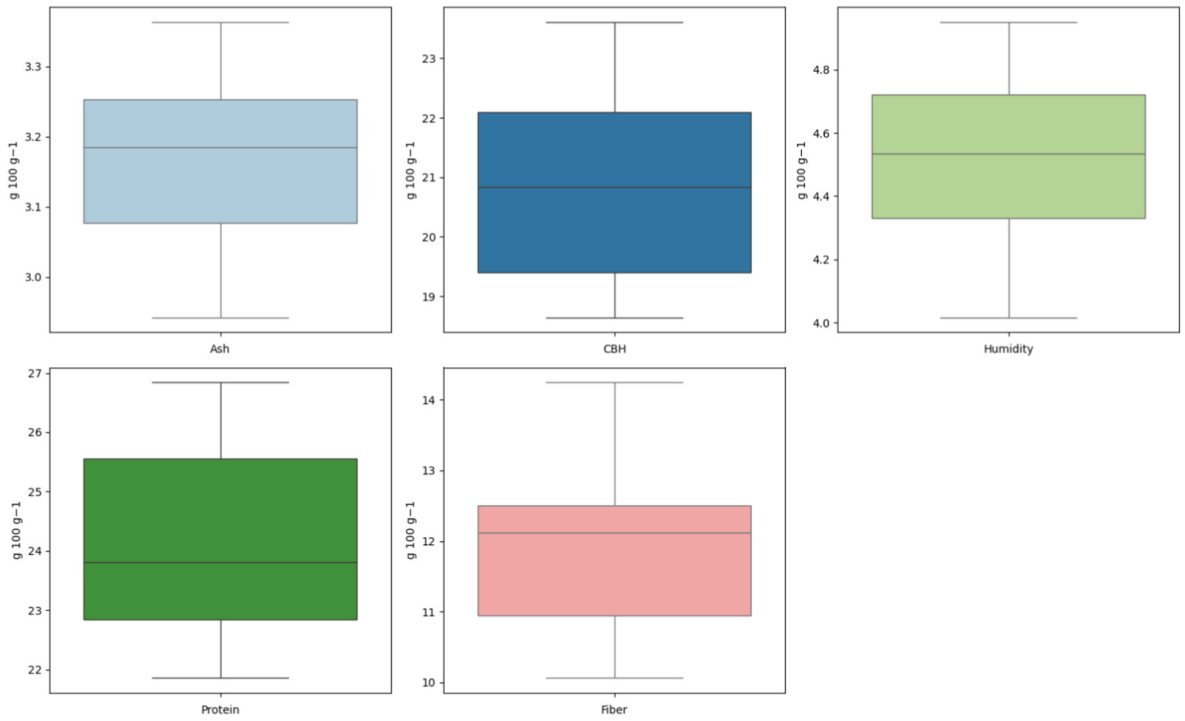

Fig. 4. Boxplots of the ash, CBH, humidity, protein and fiber content in the pistachios.

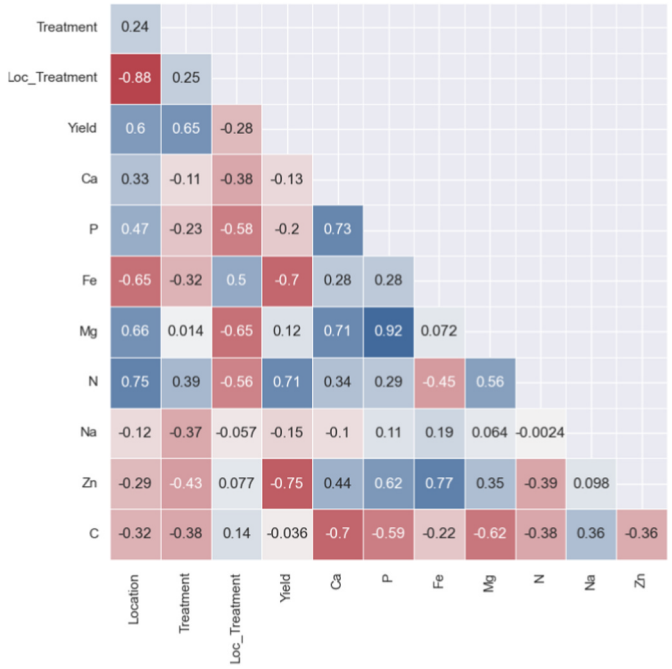

Fig. 5. Pearson correlation matrix of the nutrients and location, water treatment and yield.

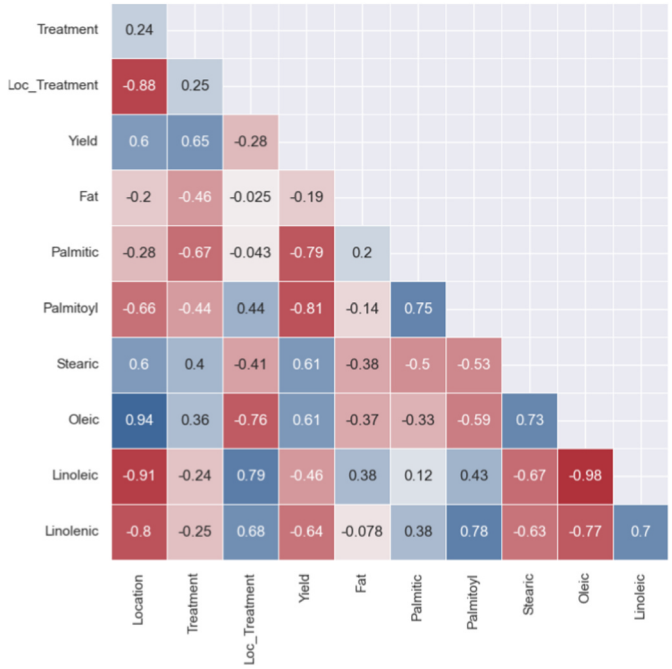

Fig. 6. Pearson correlation matrix of the fats and location, water treatment and yield.

how these biochemical factors and agronomic parameters relate to each other. The strong positive correlations between N and productivity measures underscore the importance of adequate fertilization and water availability for robust tree growth (ÜNLÜ et al., 2005). Meanwhile, Fe and Zn correlated inversely with yield, echoing prior observations that elevated concentrations of certain micronutrients do not necessarily translate to higher productivity (Liu et al., 2014). Also, the negative associations found between certain fatty acids (e.g., oleic vs. linoleic) highlighted the complex lipid dynamics in pistachios, where genetic, climatic, and irrigation variables can tip fatty acid synthesis toward one pathway or another (Kurasiak-Popowska et al., 2021; Tomé-Rodríguez et al., 2023).

Regarding predictive modelling, the PLS consistently demonstrated higher accuracy and lower errors than SVM and ETR for most parameters. This suggests that linear approaches effectively capture the dominant spectral features linked to nutrient or fat composition, making PLS particularly suited for these types of agricultural datasets. Nonetheless, SVM still performed competitively for several parameters, such as Zn

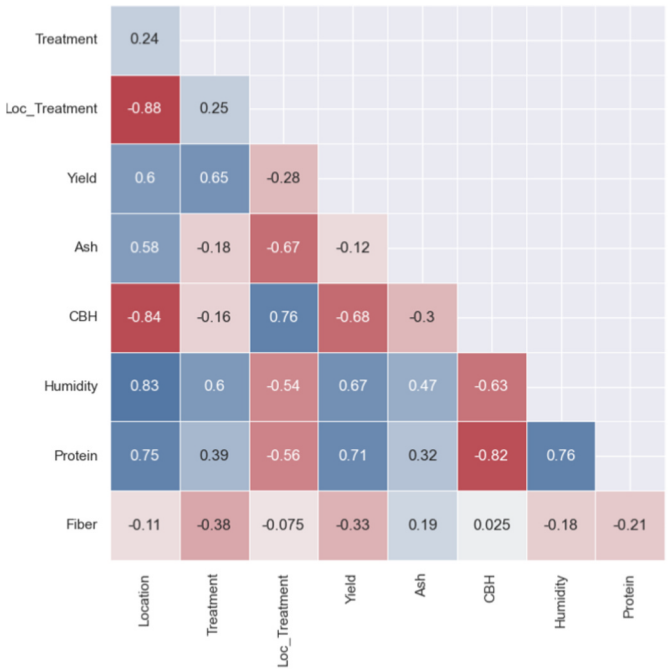

Fig. 7. Pearson correlation matrix of ash, CBH, humidity, protein and fiber, location, water treatment, and yield.

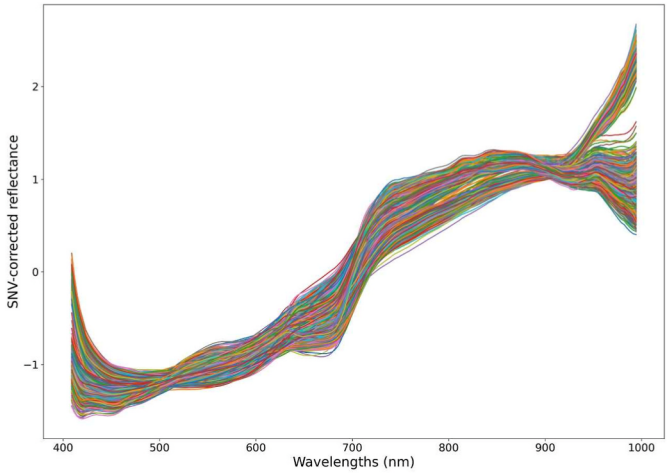

Fig. 8. All spectra with SNV scatter correction.

and oleic acid, an outcome mirrored in studies using advanced ML for compositional analysis in nuts and seeds (Hu et al., 2021; Tahmasbian et al., 2021). ETR excelled in a few scenarios (e.g., some minerals), suggesting that tree-based models may capture complex, non-linear patterns not fully addressed by PLS in specific cases.

Despite these promising results, fibre content stood out as difficult to predict across all tested models. The weak correlations observed in the dataset may indicate that fibre biosynthesis in pistachios is governed by factors unrelated, or only marginally related to the spectral signals captured in the 400–1000 nm range. This aligns with other findings where certain plant constituents, especially structural components like lignin or cellulose, can be less responsive to standard HSI ranges, as in grassland, where (Zhao et al., 2024) obtained R^2 values of 0.58.

Overall, the application of HSI combined with ML proved to be a powerful, non-destructive tool for assessing key pistachio quality traits. By pinpointing the spectral fingerprints linked to irrigation regimes and local growing conditions, our approach offers a rapid means to optimize orchard management strategies, particularly in regions facing water scarcity or in need of precise nutrient management. Such insights build on previous classification results (Martínez-Peña et al., 2024) by providing concrete numerical predictions that producers can use to fine-tune irrigation schedules and fertilizer programs. Importantly, the present models should still be regarded as preliminary: they were trained on a single season, two locations and one cultivar, using 2818 spectra, and their robustness would benefit from future datasets including multiple years, additional orchards and different cultivars to better capture environmental and genetic variability. The spectral range considered (400–1000 nm) corresponds to the sensitivity of the portable VIS–NIR camera used, which, despite being relatively low cost and field deployable, mainly captures indirect signatures of composition through pigment, water and generic organic functional group absorptions (O–H, C–H, C–O). While this is a wavelength region widely used to model moisture, protein, carbohydrates, lipids and associated minerals in agri food products, more specific and direct responses would likely be obtained with extended NIR/SWIR systems. Thus, the current results demonstrate that a portable and user friendly device can already deliver useful predictions at relatively low cost, but future work should critically extend and stress test the models using independent external datasets (different years and sites) and, where feasible, higher range NIR/SWIR cameras.

In terms of predictive modelling, the study compared the performance of PLS, SVM, and ETR. PLS emerged as the most accurate model overall, closely followed by SVM. Specific parameters such as N and Zn content were predicted with high accuracy, achieving R^2 values of 0.75 and 0.81, respectively. Similarly, predictions for fatty acids, including palmitic, palmitoleic, oleic, and linoleic acids, yielded R^2 values of 0.82, 0.82, 0.91, and 0.87, respectively. Other parameters like ash content, CBH and humidity also showed strong predictive performance with R^2 scores of 0.81, 0.87, and 0.84. While these results are encouraging, a critical evaluation highlights that generalization outside the current experimental domain remains to be demonstrated.

5. Conclusion

The findings of this study emphasize the deep influence of environmental conditions, agronomic practices, and their interactions on the composition and quality of pistachios. This study reveals how different geographic locations and irrigation regimes affect key quality traits while also evaluating the predictive performance of various ML models. The study highlights that location plays a pivotal role in determining pistachio quality. For instance, samples from Moraleja de las Panaderas consistently exhibited higher concentrations of nutrients such as N and P, along with high levels of oleic acid, protein, and ash content. Irrigation regimes also showed a significant impact, with high-water treatments leading to increased levels of key nutrients, saturated fatty acids, and ash content compared to control-water treatments. Furthermore, the interaction between location and irrigation treatment underscored the complexity of these relationships.

In terms of predictive modelling, the study compared the performance of PLS, SVM, and ETR. PLS emerged as the most accurate model overall, closely followed by SVM. Specific parameters such as N and Zn content were predicted with high accuracy, achieving R^2 values of 0.75 and 0.81, respectively. Similarly, predictions for fatty acids, including palmitic, palmitoleic, oleic, and linoleic acids, yielded R^2 values of 0.82, 0.82, 0.91, and 0.87, respectively. Other parameters like ash content, CBH and humidity also showed strong predictive performance with R^2 scores of 0.81, 0.87, and 0.84.

These findings underscore the potential of HSI as a rapid and non-destructive technique for assessing the nutritional composition and fatty acid profiles of pistachios under varying irrigation regimes. However, further research is needed to explore the underlying mechanisms driving these observed relationships and to assess their broader implications for production strategies and product quality.

Table 5
Classification metrics of the models for the pistachio nutrient predictions.

	Model	Metric	Ca	P	Fe	Mg	N	Na	Zn	C
Train	**PLS**	R^2	0.470	0.710	0.660	0.640	0.760	0.220	0.810	0.580
		MAE	0.049	0.020	1.600	0.004	0.090	2.660	0.820	0.260
		RMSE	0.006	0.028	2.050	0.005	0.120	3.240	1.120	0.320
	SVM	R^2	0.460	0.660	0.660	0.590	0.750	0.220	0.820	0.640
		MAE	0.005	0.018	1.400	0.003	0.088	2.360	0.700	0.210
		RMSE	0.006	0.030	2.080	0.005	0.120	3.240	1.080	0.290
	ETR	R^2	0.610	0.700	0.710	0.720	0.760	0.200	0.830	0.660
		MAE	0.004	0.020	1.390	0.003	0.096	2.860	0.650	0.230
		RMSE	0.005	0.028	1.900	0.004	0.120	3.260	1.040	0.280
Test	**PLS**	R^2	0.440	0.700	0.650	0.630	0.750	0.200	0.810	0.550
		MAE	0.005	0.020	1.540	0.004	0.092	2.610	0.820	0.260
		RMSE	0.006	0.028	2.010	0.005	0.120	3.220	1.110	0.330
	SVM	R^2	0.410	0.650	0.620	0.590	0.730	0.120	0.800	0.550
		MAE	0.005	0.020	1.440	0.003	0.095	2.470	0.790	0.250
		RMSE	0.007	0.030	2.070	0.005	0.120	3.360	1.130	0.330
	ETR	R^2	0.390	0.600	0.480	0.600	0.720	0.120	0.760	0.360
		MAE	0.005	0.022	1.840	0.003	0.085	2.730	0.760	0.310
		RMSE	0.007	0.042	2.430	0.005	0.120	3.370	1.180	0.390

Table 6
Classification metrics of the models for the pistachio fat predictions.

	Model	Metric	Palmitic	Palmitoyl	Stearic	Oleic	Linoleic	Linolenic	Fat
Train	**PLS**	R^2	0.850	0.840	0.560	0.910	0.880	0.740	0.380
		MAE	0.144	0.049	0.057	0.610	0.680	0.018	0.620
		RMSE	0.190	0.061	0.067	0.780	0.880	0.026	0.800
	SVM	R^2	0.830	0.840	0.560	0.900	0.870	0.650	0.430
		MAE	0.140	0.045	0.053	0.600	0.660	0.017	0.540
		RMSE	0.210	0.061	0.069	0.820	0.920	0.030	0.770
	ETR	R^2	0.710	0.760	0.650	0.960	0.890	0.800	0.620
		MAE	0.210	0.060	0.048	0.390	0.650	0.017	0.500
		RMSE	0.270	0.075	0.061	0.500	0.830	0.022	0.620
Test	**PLS**	R^2	0.820	0.820	0.550	0.910	0.870	0.700	0.320
		MAE	0.150	0.051	0.058	0.600	0.680	0.020	0.640
		RMSE	0.210	0.063	0.069	0.780	0.880	0.029	0.840
	SVM	R^2	0.790	0.800	0.530	0.900	0.860	0.600	0.330
		MAE	0.160	0.052	0.570	0.660	0.730	0.019	0.620
		RMSE	0.230	0.067	0.071	0.870	0.960	0.033	0.830
	ETR	R^2	0.700	0.750	0.320	0.780	0.770	0.750	0.610
		MAE	0.210	0.061	0.068	1.000	0.960	0.020	0.510
		RMSE	0.280	0.076	0.085	1.230	1.200	0.025	0.640

Table 7
Classification metrics of the models for the pistachio ash, CBH, humidity, protein and fiber predictions.

	Model	Metric	Ash	CBH	Humidity	Protein	Fiber
Train	**PLS**	R^2	0.830	0.880	0.870	0.760	0.240
		MAE	0.060	0.410	0.070	0.580	0.840
		RMSE	0.079	0.520	0.090	0.740	1.270
	SVM	R^2	0.082	0.880	0.860	0.760	0.210
		MAE	0.057	0.370	0.067	0.540	0.710
		RMSE	0.081	0.520	0.092	0.740	1.290
	ETR	R^2	0.730	0.780	0.820	0.760	0.200
		MAE	0.071	0.550	0.950	0.600	0.700
		RMSE	0.100	0.690	0.100	0.740	1.350
Test	**PLS**	R^2	0.810	0.870	0.840	0.750	0.180
		MAE	0.063	0.410	0.076	0.580	0.920
		RMSE	0.085	0.530	0.097	0.740	1.430
	SVM	R^2	0.790	0.860	0.820	0.730	0.100
		MAE	0.065	0.420	0.080	0.600	0.860
		RMSE	0.090	0.550	0.100	0.770	1.500
	ETR	R^2	0.710	0.700	0.780	0.740	0.050
		MAE	0.083	0.630	0.096	0.590	1.000
		RMSE	0.120	0.820	0.120	0.760	1.550

CRediT authorship contribution statement

Raquel Martínez-Peña: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Salvador Castillo:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis. **Sergio Vélez:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Sara Álvarez:** Writing – review & editing, Resources, Project administration, Funding acquisition.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Disclaimer

Any practical applications of this research should take into consideration site-specific conditions and further verification.

AI assistance disclaimer

This paper has been developed with the support of artificial intelligence (AI) tools, specifically OpenAI's ChatGPT, which assisted in the revision of the English language. All ideas, interpretations, and results presented in this manuscript are the sole responsibility of the authors.

Funding

This work was funded by MCIN/AEI/10.13039/501100011033 and European Union « NextGenerationEU»//PRTR, grant number RYC2021-033890. Co-financed by FEADER funds and Junta de Castilla y León (Spain).

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.

Acknowledgements

We would like to thank the members of the physicochemical laboratory of the “Instituto Tecnológico Agrario de Castilla y León” for their help and for carrying out the nutritional quality analysis of pistachio nuts. Furthermore.

- Dr. Raquel Martínez's actual contract has been supported by Regional Institute of Agri-Food and Forestry Research and Development of Castilla-La Mancha (IRIAF).
- Dr. Sergio Vélez's Distinguished Researcher contract (Beatriz Galindo BG23/00073) is funded by the Spanish Ministry of Science, Innovation and Universities under the 2023 call, within the framework of the State Plan for Scientific, Technical, and Innovation Research (PEICTI) 2021–2023.

Data availability

Data will be made available upon reasonable request.

References

Allakonon, M. G. B., Zakari, S., Tovihoudji, P. G., Fatondji, A. S., & Akponikpè, P. B. I. (2022). Grain yield, actual evapotranspiration and water productivity responses of maize crop to deficit irrigation: A global meta-analysis. *Agricultural Water Management, 270*, Article 107746. https://doi.org/10.1016/J.AGWAT.2022.107746

Alvarez-Vanhard, E., Corpetti, T., & Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing, 3*, Article 100019. https://doi.org/10.1016/J.SRS.2021.100019

Anjum, M. N., Cheema, M. J. M., Hussain, F., & Wu, R. S. (2023). Precision irrigation: Challenges and opportunities. *Precision Agriculture: Evolution, Insights and Emerging Trends*, 85–101. https://doi.org/10.1016/B978-0-443-18953-1.00007-6

Bonifazi, G., Capobianco, G., Gasbarrone, R., & Serranti, S. (2021). Contaminant detection in pistachio nuts by different classification methods applied to short-wave infrared hyperspectral images. *Food Control, 130*, Article 108202. https://doi.org/10.1016/J.FOODCONT.2021.108202

Fader, M., Shi, S., Von Bloh, W., Bondeau, A., & Cramer, W. (2016). Mediterranean irrigation under climate change: More efficient irrigation needed to compensate for increases in irrigation water requirements. *Hydrology and Earth System Sciences, 20* (2), 953–973. https://doi.org/10.5194/HESS-20-953-2016

Gilman, K. L. (2021). Pistachio yields and nut quality determination and the relationship between soil characteristics. https://scholarworks.calstate.edu/concern/theses/4f16c8375.

Goldhamer, D. A., & Beede, R. H. (2004). Regulated deficit irrigation effects on yield, nut quality and water-use efficiency of mature pistachio trees. *The Journal of Horticultural Science and Biotechnology, 79*(4), 538–545. https://doi.org/10.1080/14620316.2004.11511802

Hu, N., Li, W., Du, C., Zhang, Z., Gao, Y., Sun, Z., Yang, L., Yu, K., Zhang, Y., & Wang, Z. (2021). Predicting micronutrients of wheat using hyperspectral imaging. *Food Chemistry, 343*, Article 128473. https://doi.org/10.1016/J.FOODCHEM.2020.128473

Kanber, R., Yazar, A., Önder, S., & Köksal, H. (1993). Irrigation response of Pistachio (Pistacia vera L.). *Irrigation Science, 14*(1), 7–14. https://doi.org/10.1007/BF00195000/METRICS

Khadivi, A., Nikoogoftar-Sedghi, M., & Tunç, Y. (2025). Agronomic characteristics, mineral nutrient content, antioxidant capacity, biochemical composition, and fatty acid profile of Iranian pistachio (Pistacia vera L.) cultivars. *BMC Plant Biology, 25*(1), 68. https://doi.org/10.1186/S12870-025-06094-9/FIGURES/5

Kudoyarova, G. R., Dodd, I. C., Veselov, D. S., Rothwell, S. A., & Yu Veselov, S. (2015). Common and specific responses to availability of mineral nutrients and water. *Journal of Experimental Botany, 66*(8), 2133–2144. https://doi.org/10.1093/JXB/ERV017

Kurasiak-Popowska, D., Graczyk, M., Przybylska-Balcerek, A., & Stuper-Szablewska, K. (2021). Influence of variety and weather conditions on fatty acid composition of winter and spring Camelina sativa varieties in Poland. *European Food Research and Technology, 247*(2), 465–473. https://doi.org/10.1007/S00217-020-03639-0/FIGURES/4

Li, Y., Zhao, H., Zhong, T., Wu, H., Peng, W., Li, Z., Huang, W., Huang, S., Shi, B., & Li, Y. (2024). Temperature and precipitation synergistically affect yield, harvesting time and post-processing quality of tropical macadamia nuts. *Atmosphere, 15*(11), 1299. https://doi.org/10.3390/ATMOS15111299, 2024, Vol. 15, Page 1299.

Liu, H., Wang, Z. H., Li, F., Li, K., Yang, N., Yang, Y., Huang, D., Liang, D., Zhao, H., Mao, H., Liu, J., & Qiu, W. (2014). Grain iron and zinc concentrations of wheat and their relationships to yield in major wheat production areas in China. *Field Crops Research, 156*, 151–160. https://doi.org/10.1016/J.FCR.2013.11.011

Martínez-Peña, R., Castillo-Gironés, S., Álvarez, S., & Vélez, S. (2024). Tracing pistachio nuts' origin and irrigation practices through hyperspectral imaging. *Current Research in Food Science, 9*, Article 100835. https://doi.org/10.1016/J.CRFS.2024.100835

Martínez-Peña, R., Vélez, S., Vacas, R., Martín, H., & Álvarez, S. (2023). Remote sensing for sustainable pistachio cultivation and improved quality traits evaluation through thermal and non-thermal UAV vegetation indices. https://doi.org/10.3390/app13137716.

Memmi, H., Gijón, M. C., Couceiro, J. F., & Pérez-López, D. (2016). Water stress thresholds for regulated deficit irrigation in pistachio trees: Rootstock influence and effects on yield quality. *Agricultural Water Management, 164*, 58–72. https://doi.org/10.1016/J.AGWAT.2015.08.006

Mohammadi-Moghaddam, T., Razavi, S. M. A., Taghizadeh, M., Pradhan, B., Sazgarnia, A., & Shaker-Ardekani, A. (2018). Hyperspectral imaging as an effective tool for prediction the moisture content and textural characteristics of roasted pistachio kernels. *Journal of Food Measurement and Characterization, 12*(3), 1493–1502. https://doi.org/10.1007/S11694-018-9764-X/TABLES/4

Nielsen, S. S. (2010). Food analysis laboratory manual. In S. S. Nielsen (Ed.), *Food science texts series*. Boston, MA, USA: Springer US. ISBN 978-1-4419-1462-0.

Olson, D., & Anderson, J. (2021). Review on unmanned aerial vehicles, remote sensors, imagery processing, and their applications in agriculture. *Agronomy Journal, 113*(2), 971–992. https://doi.org/10.1002/AGJ2.20595

Otsu, N. (1979). Threshold SELECTION METHOD FROM GRAY-level HISTOGRAMS. *IEEE Trans Syst Man Cybern, SMC, 9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Rockström, J., & Barron, J. (2007). Water productivity in rainfed systems: Overview of challenges and analysis of opportunities in water scarcity prone savannahs. *Irrigation Science, 25*(3), 299–311. https://doi.org/10.1007/S00271-007-0062-3/FIGURES/10

Rouphael, Y., Cardarelli, M., Schwarz, D., Franken, P., & Colla, G. (2012). Effects of drought on nutrient uptake and assimilation in vegetable crops. *Plant Responses to Drought Stress: From Morphological to Molecular Features, 9783642326530*, 171–195. https://doi.org/10.1007/978-3-642-32653-0_7

Saha, D., & Manickavasagan, A. (2021). Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Current Research in Food Science, 4*, 28–44. https://doi.org/10.1016/J.CRFS.2021.01.002

Sedaghati, N., & Hokmabadi, H. (2015). Optimizing pistachio irrigation management using the relationship between echo-physiological characteristics and water stress. *Journal of Agricultural Science and Technology A, 17*(1), 189–200. http://jast.modares.ac.ir/article-23-2572-en.html.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*(3–4), 125–161. https://doi.org/10.1016/J.EARSCIREV.2010.02.004

Sun, D. W. (2010). Hyperspectral imaging for food quality analysis and control. *Hyperspectral Imaging for Food Quality Analysis and Control, 1–477*. https://doi.org/10.1016/C2009-0-01853-4

Tahmasbian, I., Wallace, H. M., Gama, T., & Hosseini Bai, S. (2021). An automated non-destructive prediction of peroxide value and free fatty acid level in mixed nut samples. *LWT, 143*, Article 110893. https://doi.org/10.1016/J.LWT.2021.110893

Tomé-Rodríguez, S., Barba-Palomeque, F., Ledesma-Escobar, C. A., Miho, H., Díez, C. M., & Priego-Capote, F. (2023). Influence of genetic and interannual factors on the fatty acids profile of virgin olive oil. *Food Chemistry, 422*, Article 136175. https://doi.org/10.1016/J.FOODCHEM.2023.136175

Ünlü, M., Kanber, R., Steduto, P., Aydin, Y., & Diker, K. (2005). Effects of different water and nitrogen levels on the yield and periodicity of pistachio (Pistacia vera L.). *Turkish Journal of Agriculture and Forestry, 29*(1), 39–49.

Wang, S., & Zhang, Z. (2011). Effects of climate change on water resources in China. *Climate Research, 47*(1–2), 77–82. https://www.int-res.com/abstracts/cr/v47/n1-2/p77-82/.

Zhao, Y., Xu, D., Li, S., Tang, K., Yu, H., Yan, R., Li, Z., Wang, X., & Xin, X. (2024). Comparative analysis of feature importance algorithms for Grassland aboveground biomass and nutrient prediction using hyperspectral data. *Agriculture, 14*(3), 389. https://doi.org/10.3390/AGRICULTURE14030389, 2024, Vol. 14, Page 389.

https://dspace.org/

HSI and ML for non-destructive pistachio quality assessment: influence of location and irrigation on nutrient and fat composition

Martínez-Peña, Raquel; Castillo-Gironés, Salvador; Vélez, Sergio; Álvarez, Sara

2026-06
Elsevier
https://doi.org/10.1016/j.foodcont.2026.112011
Martínez-Peña R, Castillo-Gironés S, Vélez S, Álvarez S. (2026) HSI and ML for non-destructive pistachio quality assessment: influence of location and irrigation on nutrient and fat composition. Food Control, Volume 184, June 2026, Article number 112011

Downloaded from a DSpace repository. https://dspace.org/