



# Non-destructive quality estimation and internal pigmentation discrimination of ‘Sanguinelli’ blood oranges using hyperspectral imaging

Salvador Castillo-Girones<sup>a</sup> , María E. García-Pastor<sup>b</sup> , María Serrano<sup>b</sup> , Daniel Valero<sup>c</sup> , Sergio Cubero<sup>d</sup> , Salvador Castillo<sup>c,\*</sup>

<sup>a</sup> Faculty of Engineering and Applied Sciences, Cranfield University, College Rd, Wharley End, Bedford, MK43 0AL, United Kingdom

<sup>b</sup> Department of Applied Biology, CIAGRO, University Miguel Hernández, (Alicante), Orihuela, Spain

<sup>c</sup> Department of Food Technology, CIAGRO, University Miguel Hernández, (Alicante), Orihuela, Spain

<sup>d</sup> Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10, 7, Valencia, 46113, Moncada, Spain

## ARTICLE INFO

### Keywords:

Anthocyanins

*Citrus sinensis* L. (Osbeck)

Internal pigmentation discrimination

Machine learning

Total soluble solids

## ABSTRACT

Blood oranges are gaining attention due to their vibrant red pigmentation and associated health benefits, primarily attributed to their high content of anthocyanins. These pigments not only enhance the fruit's visual appeal but also offer potent antioxidant and anti-inflammatory properties, which contribute to the prevention of metabolic disorders such as cardiovascular diseases and diabetes. As consumer demand increase, it is crucial for packing houses and breeding programs to non-destructively assess internal fruit quality parameters, including total soluble solids (TSS), titratable acidity (TA), maturity index (MI) and anthocyanin content. Understanding how climatic conditions influence these parameters is also essential for optimizing fruit quality and meeting market expectations. This study first investigates how chilling temperatures affect anthocyanin biosynthesis in ‘Sanguinelli’ blood oranges. It then aims to develop and compare models for the non-destructive prediction of key quality parameters, including TSS, total anthocyanin content and MI, using NIR hyperspectral imaging. Results indicated that early-season chilling hours were strongly correlated with increased anthocyanin accumulation. In contrast, late-season chilling was associated with a slight reduction in TSS content. Three machine learning models were employed for predictive modeling: Partial least squares (PLS), support vector machine (SVM), and a multi-layer perceptron (MLP). The prediction of TSS and anthocyanin content achieved  $R^2$  scores of 0.72 and 0.61, respectively. However, the predictions for TA and MI were not satisfactory. When assessing internal pigmentation, the best model, the MLP achieved accuracies of 79 % and 66 % for two and three different pigmentation groups, respectively. The MLP also achieved high accuracy for internal pigmentation discrimination and demonstrated good predictive capabilities for TSS and internal pigmentation. These findings confirm the potential of NIR as a rapid and reliable tool for the non-destructive quality control and internal pigmentation sorting of ‘Sanguinelli’ blood oranges.

## 1. Introduction

Oranges, particularly blood oranges [*Citrus sinensis* L. (Osbeck)], play a significant role in Europe and Spain due to their cultural, economic, and nutritional importance. The Mediterranean region leads orange production in the EU, with Spain producing 56 % of the total, followed by Italy with 24 % and Greece with 14 % (Eurostat, 2025). Blood oranges are widely appreciated by consumers, and being acknowledged as a functional food with remarkable health benefits (Pio Ávila et al., 2022). Their distinctive red colour, caused by anthocyanins, makes

them visually appealing and emphasizes their unique biochemical composition (Garrido-Auñón, Padilla-González, et al., 2025). Anthocyanins are powerful antioxidants that give blood oranges their distinctive pigmentation. These compounds have radical-scavenging properties that contribute to their anti-cancer, anti-inflammatory, and cardiovascular protective effects. Blood oranges also contain other bioactive compounds, such as flavonoids, carotenoids, ascorbic acid, and hydroxycinnamic acids, which enhance their health benefits further (Grosso et al., 2013). These health benefits are a key factor in the growing popularity of blood oranges (Lo Piero, 2015).

\* Corresponding author.

E-mail address: [scastillo@umh.es](mailto:scastillo@umh.es) (S. Castillo).

<https://doi.org/10.1016/j.foodcont.2026.112324>

Received 24 March 2026; Received in revised form 11 May 2026; Accepted 20 May 2026

Available online 24 May 2026

0956-7135/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The acceptance of blood oranges among consumers varies by country and is influenced by factors such as familiarity, sensory characteristics, and perceived health benefits. A cross-cultural study conducted in China, Mexico, Spain, and Italy by Giménez-Sanchis, Zhong, et al. (2022) found that Italians had the highest availability and consumption of blood oranges. Across all the countries studied, the main motivations for consumption were 'liking' and 'health benefits'. Spanish consumers particularly appreciated the intense pigmentation of blood oranges but found them less convenient to eat than other citrus fruits. A similar study involving Italian and Spanish consumers by the same author, Giménez-Sanchis, Pons-Gómez, et al. (2022), p. 3058, revealed that Spanish consumers preferred blood orange cultivars with less intense pigmentation to deeply red ones. In Spain specifically, consumer response to blood oranges is positively influenced by health-related claims; however, this did not significantly change the preferences of Italian consumers.

The internal quality and pigmentation of blood oranges can vary significantly between seasons, primarily due to environmental factors such as temperature. Cold temperatures are known to increase the amount of anthocyanin content in blood oranges by triggering specific genetic and biochemical pathways (Crifó et al., 2011). This process is mediated by the activation of the 'Ruby' gene, which is essential for anthocyanin biosynthesis and cold-dependent (Caruso et al., 2016; Chen et al., 2023). During colder periods, transcriptomic changes activate flavonoid biosynthesis pathways, increasing pigmentation in pulp and rind. Seasonal differences also affect other internal quality parameters, such as sugar content and acidity, which are influenced by maturation stages and environmental conditions (Caruso et al., 2016). In addition, deficit irrigation can also lead to increased total soluble solids (TSS), titratable acidity (TA) and anthocyanins (Pérez-Pérez et al., 2023). Given the importance of internal quality in blood oranges, it is crucial to determine their quality in a non-destructive way. While traditional methods require the destruction of the sample, hyperspectral imaging (HSI) is a powerful, non-destructive technology that combines imaging and spectroscopy to assess fruit quality (Cai et al., 2025; Riccioli et al., 2021; Sun, 2010). Sugar content, acidity, anthocyanins and internal pigmentation are all vital factors in determining fruit quality, and HSI could be an efficient tool for quality control in commercial operations. While the research conducted by Riccioli et al. (2021) and Cai et al. (2025) demonstrates the efficacy of HSI for evaluating quality in blond oranges, these studies do not specifically focus on blood oranges. Consequently, a knowledge gap exists regarding the use of these non-destructive technologies to estimate the unique internal attributes of blood oranges without fruit destruction.

One major benefit of HSI is its ability to analyse multiple internal parameters simultaneously without damaging the fruit. However, the thick rind of citrus fruits poses a challenge as it limits light penetration since light can only penetrate a few millimetres into the rind as proved in persimmon fruit by Rodríguez-Ortega et al. (2023), which reduces the accuracy of prediction of pulp parameter, such as pigmentation, sugar content or acidity.

To extract meaningful information from the high-dimensional spectral data generated by HSI, exploratory analysis and supervised modelling approaches are commonly employed. Principal Component Analysis (PCA) is widely used as an exploratory step to identify dominant sources of variance, and reveal clustering patterns among samples prior to model development (Duntelman, 1989; Greenacre et al., 2022). Building on this foundation, supervised regression and classification methods can then be applied to establish predictive relationships between spectral features and quality parameters. Among these, Partial Least Squares regression (PLS) has been extensively adopted in food quality assessment due to its ability to handle collinear spectral data and small sample sizes. Support Vector Machines (SVM) offer robust performance in both regression and classification tasks by identifying optimal decision boundaries in high-dimensional feature spaces, making them well-suited for complex spectral datasets. More recently, Multilayer Perceptron

networks (MLP) have demonstrated strong predictive capacity for non-linear relationships in spectroscopic applications (Cen & He, 2007; Müller & Guido, 2016; Sharma et al., 2021). The combined use of these methods provides a comprehensive analytical framework for the non-destructive assessment of fruit quality from HSI data.

The present research study examines the viability of HSI for these complex measurements. Several studies have previously examined the internal quality of citrus fruits using hyperspectral imaging or spectroscopy, with some focusing particularly on the impact of environmental factors on peaches, sweet lemon or orange quality parameters (Moomkesh et al., 2017; Pan et al., 2016; Riccioli et al., 2021; Tian et al., 2022). To date, no studies have specifically applied HSI to assess quality in 'Sanguinelli' blood oranges. Therefore, this research aims to non-destructively predict key parameters, including TSS, TA, anthocyanin content, and maturity index (MI). Additionally, given the significance of anthocyanins for human health and the distinctive internal pigmentation of blood oranges, which enhances their value compared to blond oranges, this study also emphasizes in predicting and classifying blood oranges based on their internal pigmentation which would be very useful for both packing houses and retailers.

## 2. Materials and methods

### 2.1. Plant material and climate data

Blood orange trees [*Citrus sinensis* L. (Osbeck)] of the 'Sanguinelli' cultivar were grafted onto *Citrus macrophylla* vigorous rootstock in 2011, were used in the experiments, which were conducted over three seasons (2022–2023, 2023–2024 and 2024–2025) when the trees were 10, 11 and 12 years old. The research was carried out in a commercial plot owned by Las Moreras Fruits & Veggies and located in Algorfa, Alicante, Spain (GPS coordinates: 38.08369, −0.77392). A total of 1080 'Sanguinelli' blood oranges were used in this study, which was conducted over three different seasons in order to capture sufficient variability. This is because fruit quality is closely linked to climatic conditions, which can vary significantly from year to year. A total of 360 oranges were analysed in each season: 2023 (25th January), 2024 (7th February), and 2025 (14th January). Samples were collected at random from the orchards in January and February to ensure variability.

Climate data were obtained from the weather station located at Rojales (Alicante, Spain), named 'El Molino' from the (AEMET - Estación meteorológica Rojales, n.d.).

### 2.2. Hyperspectral imaging system and data acquisition

Each fruit was independently imaged using a hyperspectral system at the Valencian Institute for Agricultural Research (IVIA) in Spain. This consisted of a NIR hyperspectral camera (Headwall NIR R-Series camera, Bolton, MA, USA) with a spectral range of 900–1700 nm and 5 nm spectral resolution; a Kowa lens (Kowa LM25HC-SW, Nagoya, Japan); a dark box; an illumination dome containing 18 halogen lamps of 55 W and 12 V (LLB477 Lucas Lighting, Itasca, IL USA); two fans; a conveyor belt travelling at 20 mm s<sup>−1</sup>; and a computer. The reflectance images were then corrected using white and dark references (Spectralon, NH, USA) and subsequently segmented.

Segmentation was performed by applying PCA exclusively as a pre-processing tool to facilitate background removal from the hyperspectral images. This use of PCA is strictly instrumental, serving as an image processing step. Specifically, the first principal component (PC1) was extracted from each hyperspectral image using the scikit-learn Python library, as it captured the greatest variance in the data and provided sufficient contrast between the sample and the background. This contrast enabled straightforward binarization and mask generation using Otsu's thresholding method from the scikit-image library. The resulting mask was then applied to the original hyperspectral image to isolate the region of interest and remove the background. Subsequently,

the mean spectrum of each segmented image was extracted and saved, forming the spectral dataset for each year. Finally, Standard Normal Variate (SNV) scatter correction was applied to the raw reflectance spectra. SNV is a spectral preprocessing technique that standardizes each individual spectrum by subtracting its mean and dividing by its standard deviation, thereby reducing the effects of scatter and baseline variations caused by differences in surface properties or illumination (Fearn et al., 2009).

### 2.3. Quality analysis

After the spectral data acquisition, blood orange juice was extracted from each fruit individually using an electric juicer. All destructive analyses were conducted within the CIAGRO postharvest facilities at Miguel Hernández University (Orihuela, Alicante, Spain). The TSS content was measured in duplicate for each individual juice sample using a digital refractometer (Hanna Instruments, Rhode Island, USA) at 20 °C, with results (mean  $\pm$  SE) expressed as ° Brix. On the other hand, the TA was also determined in duplicate per fruit via automatic titration (1010 OMNIS Sample Robot S, Metrohm, Herisau, Switzerland) to a pH endpoint of 8.10, using 1 mL of juice in 25 mL of distilled water and 0.1 N NaOH. TA results (mean  $\pm$  SE) were reported as percentage (%) of citric acid in fresh weight (FW). Finally, the MI of the blood oranges was shown as an absolute value obtained by dividing the TSS by the TA (Serna-Escolano, Giménez, Serrano, et al., 2024).

Juice anthocyanin extractions and quantifications were carried out using a method adapted from that reported by (Habibi et al., 2020). The total anthocyanin content of the juice was measured by extracting 2.5 mL of juice with 5 mL of methanol/hydrochloric acid/water (25:1:24, v/v/v) using a mortar and pestle. The extracts were then centrifuged at 10,000 g for 10 min at 4 °C. Total anthocyanin content in the resulting supernatant was quantified by measuring the absorbance at 520 nm using a spectrophotometer (UV-1900 Spectrophotometer, Shimadzu USA Manufacturing Inc., Canby, USA). Quantification was performed using an external calibration curve of cyanidin-3-O-glucoside (Sigma-Aldrich, St. Louis, MO, USA) as a standard. The calibration standards were prepared in the extraction solvent across a concentration range of 0.5 to 100 mg L<sup>-1</sup>, exhibiting high linearity ( $R^2 = 0.998$ ). Results were expressed as mg of cyanidin-3-O-glucoside equivalents (C3G) per L of juice.

### 2.4. Model development and calibration

Predictive internal pigmentation groups were established by applying k-means clustering (Scikit-learn, Python) to anthocyanin values. The optimal number of clusters was determined using the Elbow method, which assesses the explained variance across increasing numbers of clusters; the point at which additional clusters yield diminishing returns ('the elbow') was selected as optimal (Madhulatha, 2012). Differences in the internal chemistry of 'Sanguinelli' blood oranges such as TSS, acidity and anthocyanin levels were observed when the fruit was grown under different climate conditions each year. These differences affect the fruit's spectral signatures. Therefore, the models may misinterpret these climate-driven spectral shifts as changes in chemical quality. A common approach to improving calibration performance in a new validation population is to add samples from the validation dataset to the calibration dataset (Cavaco et al., 2018; Guthrie et al., 2005; Pires et al., 2022), proving it solves problems when using external validation with a different harvest season while avoiding internal calibration. This provides less realistic predictions than testing with new independent data (Serna-Escolano, Giménez, Zapata, et al., 2024; Zude et al., 2008). In this study, calibration was performed using datasets from the first two seasons; 2022–2023 and 2023–2024, and was then recalibrated using 30 % of the validation dataset which corresponds to the 2024–2025 season.

Three commonly used machine learning (ML) algorithms were

implemented: PLS, SVM, and a multilayer perceptron MLP. Model training and subsequent data analysis were conducted on a workstation equipped with a 14-core Intel i9 processor (12th generation; Intel Inc., Santa Clara, CA, USA), 64 GB of DDR4 RAM, and an 8 GB GPU (NVIDIA GeForce RTX 5070 Ti; Nvidia Inc., Santa Clara, CA, USA).

Hyperparameter optimisation was performed using 10-fold cross-validation implemented via the Python scikit-learn library. For the PLS models, the optimal number of latent variables was determined using GridSearchCV to evaluate every possible value of *n*\_components from 1 up to the maximum number of available wavelengths (170 combinations). For the SVM and MLP models, RandomizedSearchCV was utilised to explore their respective parameter spaces: for the SVM, *C* = logspace(−2, 5, 200), kernel = {linear, poly, rbf, sigmoid}, gamma = {scale, auto}, and degree = {2, 3, 4, 5}, yielding 6400 combinations; for the MLP, *n*\_layers = {2, 3, 4, 5, 6, 8, 10}, first\_layer\_nodes = {200, 150, 75, 60}, and intermediate\_layer\_nodes = {200, 179, 157, 136, 114, 92, 71, 49, 27, 6}, yielding 280 combinations. A total of 6849 combinations were explored across all three models. Across all models, the final hyperparameters were selected based on the same performance criteria: the lowest root mean square error of cross-validation for regression tasks, and the highest accuracy for classification tasks. The parameters selected for the regression and classification models are shown in Tables 1 and 2, respectively.

The performance of the regression models was evaluated on both the training and validation datasets using the coefficient of determination ( $R^2$ ), mean absolute error (MAE) and root mean squared error (RMSE).  $R^2$  indicates the proportion of the variance in the dependent variable that is explained by the independent variables. MAE measures the average absolute difference between the predicted and actual values, while RMSE also measures the average absolute error between the predicted and actual values, but expresses it in the same units as the original data. Classification models were evaluated using accuracy, precision, and recall, and confusion matrices were generated for the validation set. Accuracy measures the proportion of correctly classified samples out of the total number of samples. Precision indicates how many of the samples assigned to a given class were correctly classified. Recall indicates how many of the true samples in a given class were correctly detected by the model (Ballabio et al., 2018; Sokolova & Lapalme, 2009).

### 2.5. Statistical and spectral analysis

Statgraphics Centurion 19 (Statistical Graphics Corporation, Rockville, MA, USA) was used to perform a multifactor analysis of variance (ANOVA) and a Tukey's multiple range test on the instrumental orange quality data to analyse the effect of each season. Since neither parameter followed a normal distribution, the non-parametric Kruskal–Wallis test was also employed to identify significant differences in quality parameters based on year and environmental temperature. A Pearson correlation matrix was also created using Python and the Seaborn library to study the linear relationship between the measured quality parameters and the harvest seasons. Statistically significant differences were considered at  $p < 0.05$ .

Following spectral preprocessing, PCA was applied to the spectral dataset as an exploratory multivariate analysis tool to characterise and understand the underlying spectral differences among samples according to their composition and grouping. This application of PCA was distinct from its use in Section 2.2, where it served solely as an image segmentation aid. Here, PCA was employed in its classical statistical role: an unsupervised multivariate technique that reduces the dimensionality of the data by transforming the original correlated variables into a smaller set of uncorrelated principal components, ordered according to the amount of variance they explain (Duntelman, 1989). The resulting scores plots were used to visualise the distribution and clustering of samples in the reduced spectral space, providing insight into the main sources of spectral variation across the dataset.

**Table 1**  
Summary of configurations and hyperparameters for the regression models applied to ‘Sanguinelli’ orange quality. <sup>Y</sup>

| Parameters predicted | Models               |               |          |                                 |
|----------------------|----------------------|---------------|----------|---------------------------------|
|                      | PLS                  | SVM           | MLP      |                                 |
| TSS                  | Latent variables: 18 | Kernel:       | rbf*     | Layers:                         |
|                      |                      | Gamma:        | Scale    | Neurons per layer:              |
|                      |                      | Class weight: | Balanced | Activation function:            |
|                      |                      | C:            | 12173    | Last layer activation function: |
|                      |                      |               |          | Loss function:                  |
| TA                   | Latent variables: 12 | Kernel:       | Linear   | Layers:                         |
|                      |                      | Gamma:        | Scale    | Neurons per layer:              |
|                      |                      | Class weight: | Balanced | Activation function:            |
|                      |                      | C:            | 23272    | Last layer activation function: |
|                      |                      |               |          | Loss function:                  |
| MI                   | Latent variables: 11 | Kernel:       | Linear   | Layers:                         |
|                      |                      | Gamma:        | Scale    | Neurons per layer:              |
|                      |                      | Class weight: | Balanced | Activation function:            |
|                      |                      | C:            | 23272    | Last layer activation function: |
|                      |                      |               |          | Loss function:                  |
| Total Anthocyanins   | Latent variables: 17 | Kernel:       | Poly**   | Layers:                         |
|                      |                      | Gamma:        | Scale    | Neurons per layer:              |
|                      |                      | Class weight: | Balanced | Activation function:            |
|                      |                      | C:            | 37834    | Last layer activation function: |
|                      |                      |               |          | Loss function:                  |

<sup>Y</sup> Abbreviations: PLS (partial least squares); SVM (support vector machine); MLP (multi-layer perceptron); TSS (total soluble solids); TA (titratable acidity); MI (maturity index); rbf\* (radial basis function); Poly\*\* (polynomial of degree 3); and MSE\*\*\* (mean squared error).

**Table 2**  
Summary of configurations and hyperparameters for the classification models applied to ‘Sanguinelli’ orange quality. <sup>Y</sup>

| Parameters predicted        | Models              |               |          |                                 |
|-----------------------------|---------------------|---------------|----------|---------------------------------|
|                             | PLS                 | SVM           | MLP      |                                 |
| Internal colour in 3 groups | Latent variables: 5 | Kernel:       | rbf*     | Layers:                         |
|                             |                     | Gamma:        | 0.1      | Neurons                         |
|                             |                     | Class weight: | Balanced | per layer:                      |
|                             |                     | C:            | 2.8117   | Activation function:            |
|                             |                     |               |          | Last layer activation function: |
| Internal colour in 2 groups | Latent variables: 7 | Kernel:       | rbf*     | Layers:                         |
|                             |                     | Gamma:        | 0.1      | Neurons                         |
|                             |                     | Class weight: | Balanced | per layer:                      |
|                             |                     | C:            | 0.6866   | Activation function:            |
|                             |                     |               |          | Last layer activation function: |
|                             |                     |               |          | Loss function:                  |
|                             |                     |               |          | Loss function:                  |

<sup>Y</sup> Abbreviations: PLS (partial least squares); SVM (support vector machine); MLP (multi-layer perceptron); TSS (total soluble solids); TA (titratable acidity); MI (maturity index); rbf\* (radial basis function); Poly\*\* (polynomial of degree 3); and MSE\*\*\* (mean squared error).

**Table 3**  
Evolution of chemical and functional quality parameters and statistical significance across the 2023–2025 harvesting seasons. <sup>Y</sup>

| ANOVA & Kruskal-Wallis test | TSS (° Brix)     | TA (% citric acid) | MI              | Total anthocyanins (mg L <sup>-1</sup> ) |
|-----------------------------|------------------|--------------------|-----------------|------------------------------------------|
| Seasons                     | ***              | ***                | ***             | ***                                      |
| 2022–2023                   | 11.058 ± 0.042 b | 1.155 ± 0.011 c    | 9.847 ± 0.076 a | 21.921 ± 0.588 c                         |
| 2023–2024                   | 9.556 ± 0.043 c  | 1.188 ± 0.011 b    | 8.199 ± 0.079 c | 42.618 ± 0.597 b                         |
| 2024–2025                   | 11.195 ± 0.048 a | 1.337 ± 0.012 a    | 8.509 ± 0.087 b | 50.460 ± 0.661 a                         |

<sup>Y</sup> Levels of significance are indicated by asterisks; \*\*\* denotes significance at  $p < 0.001$ , as determined by an ANOVA and Kruskal-Wallis test. According to Tukey’s least significant difference test, different lowercase letters within the same column indicate significant differences among seasons at  $p < 0.05$  for each parameter tested. Abbreviations: TSS (total soluble solids); TA (titratable acidity); and MI (maturity index).

3. Results

3.1. Correlation analysis of quality parameters

Table 3 shows significant statistical differences in TSS, TA, MI and total anthocyanin content between seasons. Samples from 2025 season showed the highest TSS, TA and anthocyanin levels, whereas the 2023 season showed the lowest TA and anthocyanin values. The average TSS value was the highest in 2025 season at 11.195 °Brix, followed by the 2023 season at 11.058 °Brix, and the lowest was in 2024 season at 9.556 °Brix. TA values were highest in 2025 season at 1.337 % citric acid, followed by 2024 season at 1.188 % and the 2023 season at 1.155 %. Regarding MI, the highest values were found in the 2023 season (9.847), followed by 2025 season (8.509) and 2024 season (8.199). In addition, significant differences in anthocyanin levels were found among years, meaning that the internal pigmentation varied considerably from year to year. The average total anthocyanin value was 50.460 mg L<sup>-1</sup> in 2025, whereas in 2023 it was 2-fold lower (21.921 mg L<sup>-1</sup>). This indicates that the flesh pigmentation of the blood oranges in 2023 was not very red, whereas in 2025 it was much redder. The anthocyanin values in 2024 were closer to those in 2025 than in 2023, with an average value of 42.618 mg L<sup>-1</sup>.

The differences mentioned earlier may be due to the environmental conditions of each year. Therefore, the temperature evolution over the two months before harvesting was evaluated, as this is key to the final quality and pigmentation of the fruit. However, the average temperatures were similar in the three years 13.0 °C in the 2022–2023 season, 13.8 °C in the 2023–2024 season, and 13.2 °C in the 2024–2025 season (Fig. 1). Nevertheless, the daily maximum and minimum temperatures

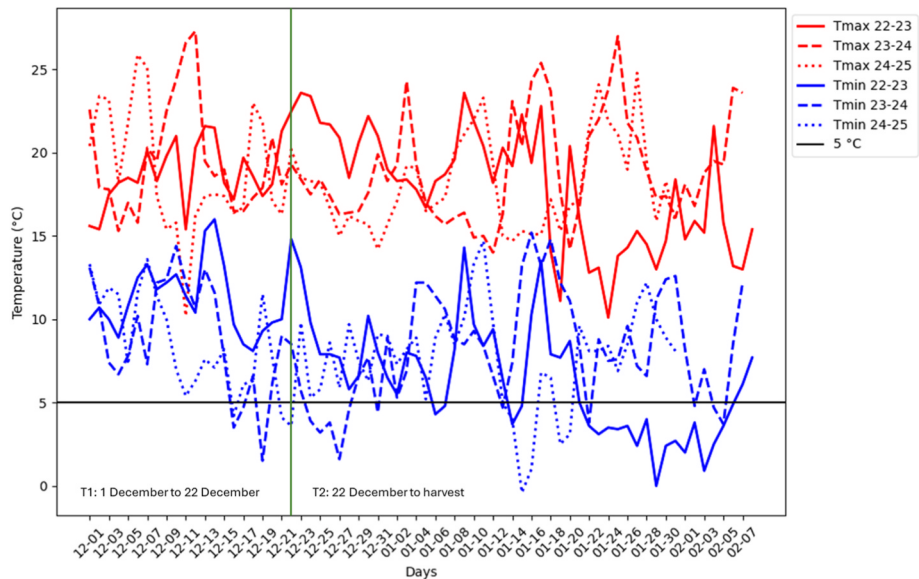


**Fig. 1.** Comparative analysis of thermal profiles: Maximum and minimum daily temperatures from December to February across three consecutive seasons (2022–2023, 2023–2024, and 2024–2025).

for the three seasons, showed apparent differences when comparing years. During the 2022–2023 season, there were no days with temperatures below 5 °C in December, and it was hotter overall, while the 2023–2024 season recorded the highest number of days with temperatures below 5 °C in December. In the 2022–2023 season, zero chilling hours (hours below 5 °C) were recorded. In the 2023–2024 season, 34 h below 5 °C were measured, and in the 2024–2025 season, 45 h below 5 °C were recorded after 22nd of January.

A Pearson correlation matrix was created to explore the relationships among various destructive parameters (TSS, TA, MI, and total anthocyanins), seasons and temperature conditions (Fig. 2). The harvesting season showed a strong correlation with anthocyanin content, with a

correlation coefficient of 0.73. To investigate the effect of temperature on anthocyanin content, the number of chilling units (hours at temperatures below 5 °C) was recorded in two periods referred to as 1st T and 2nd T. The 1st T was 1st to 22nd December, and 2nd T from 23rd December until 25th January, 7th February and 14th January for 2022–2023, 2023–2024 and 2024–2025 seasons respectively, which were the harvest dates. These periods are referred to as 1st T and 2nd T in Fig. 2.

A correlation of 0.75 was observed between anthocyanin content and the number of chilling hours recorded during the first temperature period. Samples harvested at the end of January from seasons with more chilling hours measured in December exhibited higher anthocyanin levels. For instance, in the 2024–2025 season, 45 h below 5 °C were recorded, resulting in the highest total anthocyanin levels, as shown in Table 3, with an average value of 50.460 mg L<sup>-1</sup>. In contrast, samples from the 2022–2023 season, which recorded zero chilling hours, displayed significantly lower anthocyanin concentration, with an average value of 21.921 mg L<sup>-1</sup>, as can be observed in Fig. 1 and Table 3. In the 2023–2024 season, 34 h below 5 °C were measured and the average anthocyanin content was 42.618 mg L<sup>-1</sup>. Furthermore, temperatures at the end of the season (2nd T) had a slight impact on anthocyanin content, while TSS content showed a negative correlation of –0.69. This indicates that higher chilling hours are associated with lower TSS values.

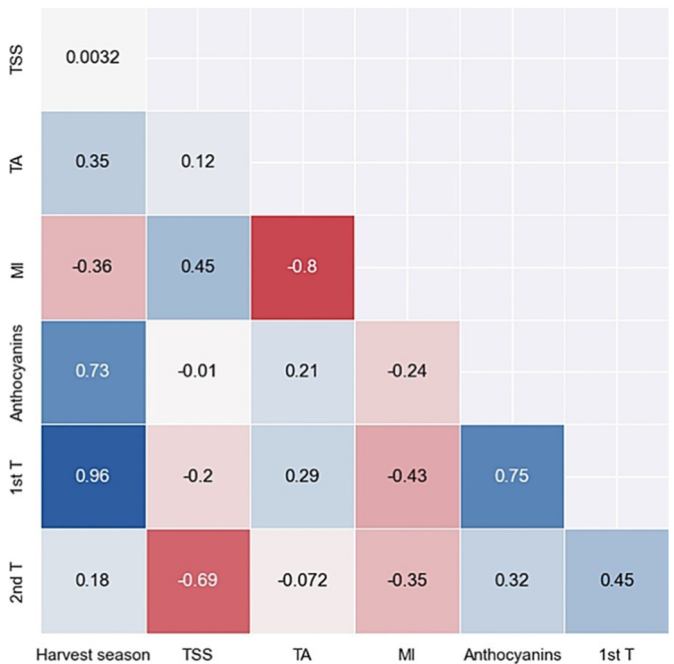


**Fig. 2.** Heatmap of Pearson correlation coefficients (*r*) among fruit quality attributes, harvest season, and thermal indicators. Abbreviations: TSS (total soluble solids); TA (titratable acidity); MI (maturity index); 1st T (from 1st to 22nd December; and 2nd T (from 23rd December until harvest dates).

3.2. Spectral exploratory analysis and internal pigmentation clustering

Packing houses need to sort oranges according to their internal pigmentation. To this end, clusters were formed according to total anthocyanin levels, since is known that the anthocyanin content, is correlated with the red pigmentation of the blood orange fruit (Garrido-Aunón, Puente-Moreno, et al., 2025). After plotting the inertias against the number of clusters, as can be observed in Fig. 3, and analysing the elbow curve, it was determined that the optimal number of clusters was two or three. Consequently, the samples were labeled according to these two methods of clustering.

In addition, Fig. 4 shows the violin plots of anthocyanin distribution for samples classified into two or three clusters. Samples with high anthocyanin levels are less common than those with low levels in both violin plots. These clusters were subsequently used to label the orange samples according to their internal pigmentation profile, enabling a spectral characterisation of each group.

The reflectance spectra of all orange samples, recorded in the

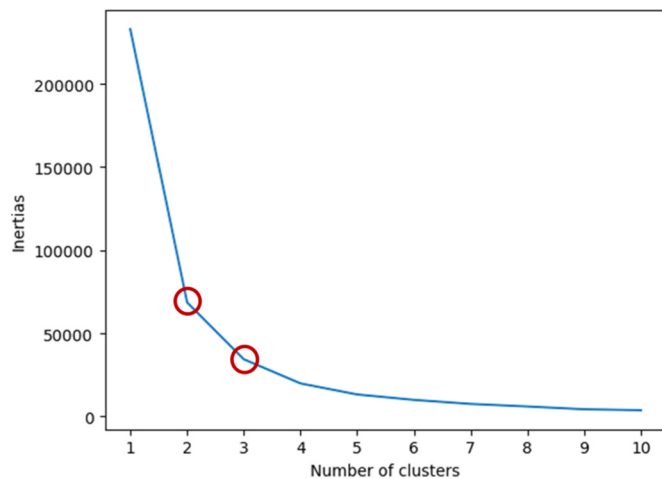


**Fig. 3.** Elbow method for determining the optimal number of clusters based on within-cluster sum of squares (Inertia).

900–1700 nm range after SNV correction, are presented in Fig. 5 left. Besides, Fig. 5 right shows the average spectra of each of the clusters representing the three pigmentation levels. Spectra followed a consistent pattern, with distinct regions exhibiting higher spectral variability: around 970 nm, associated with the second overtone of O–H stretching vibrations linked to water content, differences near 1100 nm and around 1450 nm, corresponding to the first overtone of O–H bonds. Differences in spectra were also observed between 1200 and 1300 nm, which has been associated with C–H combination bands related to carbohydrate and organic acid content in fruit tissue (Pasquini, 2003; Vranić et al.,

2020; Wu & Sun, 2013). Furthermore, separation was observed in the 1600–1700 nm region, primarily associated with the first overtone of C–H stretching vibrations, which are characteristic of soluble sugars, structural carbohydrates, and phenolic compounds. While anthocyanins do not possess isolated absorption peaks in this specific NIR range, they belong to the broader class of phenolic compounds whose aromatic C–H bonds contribute to absorption in this region (Nicolai et al., 2007). This demonstrates that SNV reduced inter-sample scatter variability, allowing subtle differences in internal chemical composition, such as those associated with anthocyanin accumulation and pigmentation levels, to become analytically accessible (Custodio-Mendoza et al., 2024).

To support the interpretation of the spectral differences observed among pigmentation groups, PCA was performed on the SNV-preprocessed spectra and the resulting scores and loadings plots (Fig. 6) were examined to identify which wavelengths are driving the group separation observed in the mean spectra.

The first five principal components accounted for 99% of the total spectral variance, confirming that the spectral dataset can be effectively summarised in a reduced dimensionality space. PC1 (52 %) displayed a near-constant loading profile across the entire spectral range, consistent with it capturing global reflectance differences driven by broadband scattering and baseline effects (Cen & He, 2007). This is corroborated by the PCA scores plot (Fig. 7), where PC1 primarily separates the High pigmentation group from Low and Medium along the horizontal axis, suggesting that high anthocyanin content is associated with a general, spectrally non-specific modulation of NIR reflectance (Elhady et al., 2024). In contrast, the higher PCs exhibited structured loading profiles linked to specific spectral regions: PC2 (24 %) showed pronounced loadings around 1500–1600 nm, coinciding with the divergence observed between groups at longer wavelengths. PC3 (13 %) displayed sharp features around 960–1000 nm and 1100–1150 nm, associated

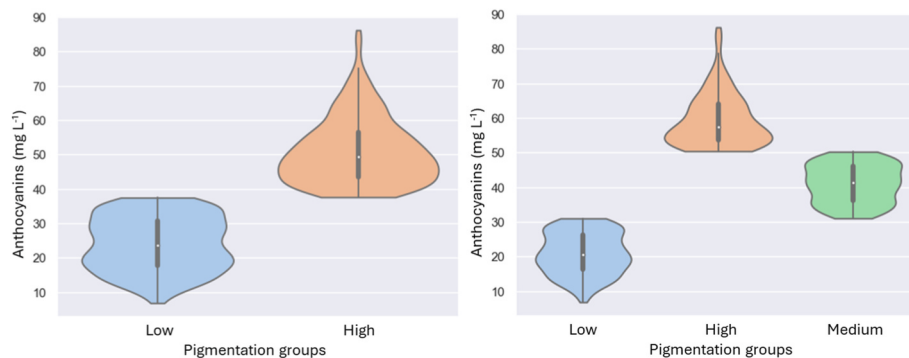


**Fig. 4.** Violin plots of total anthocyanin content distribution segmented by K-means clustering [ $k = 2$  (left) and  $k = 3$  (right)].

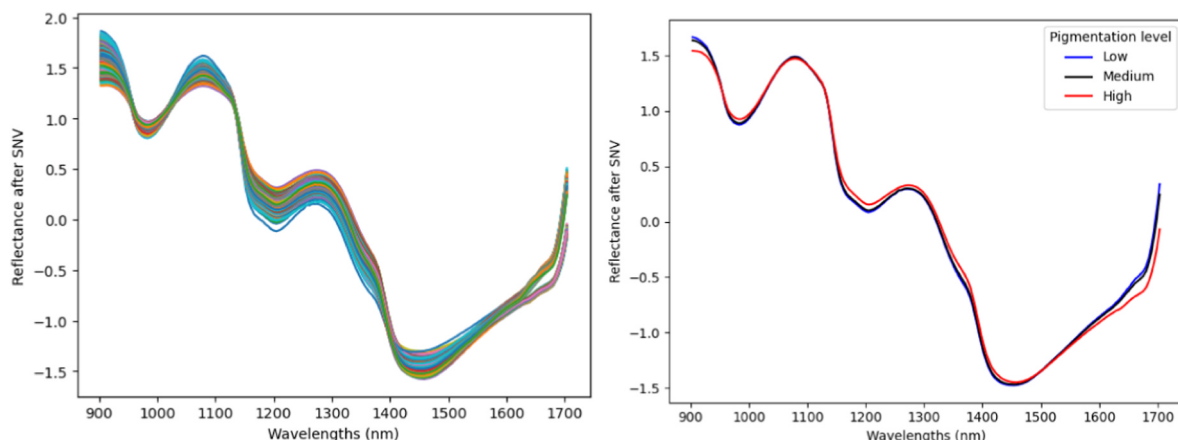


**Fig. 5.** Original spectra after SNV scatter correction (left) and average spectra after SNV correction of the three different pigmentation levels after clustering (right).

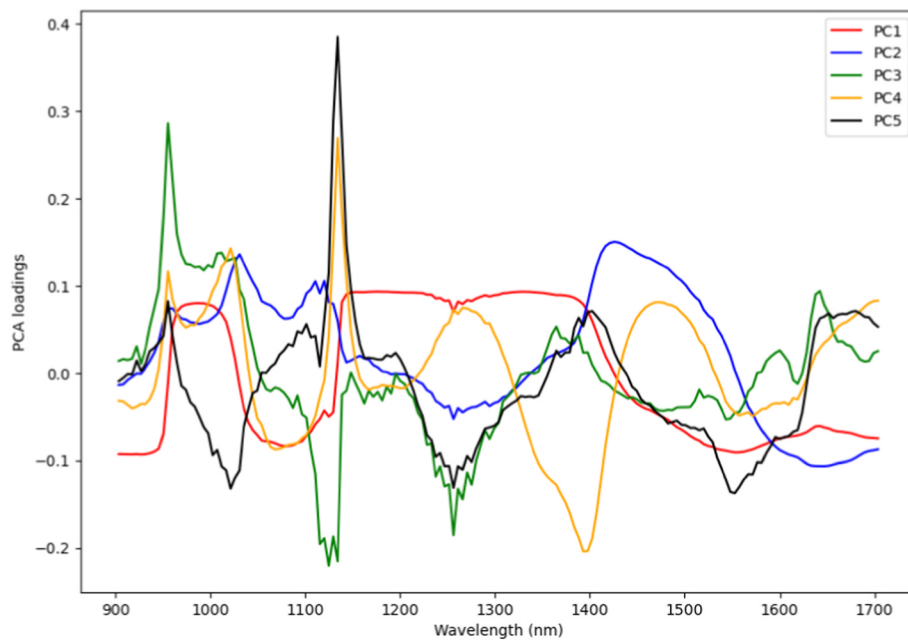


Fig. 6. PCA loadings profiles across the NIR spectral range.

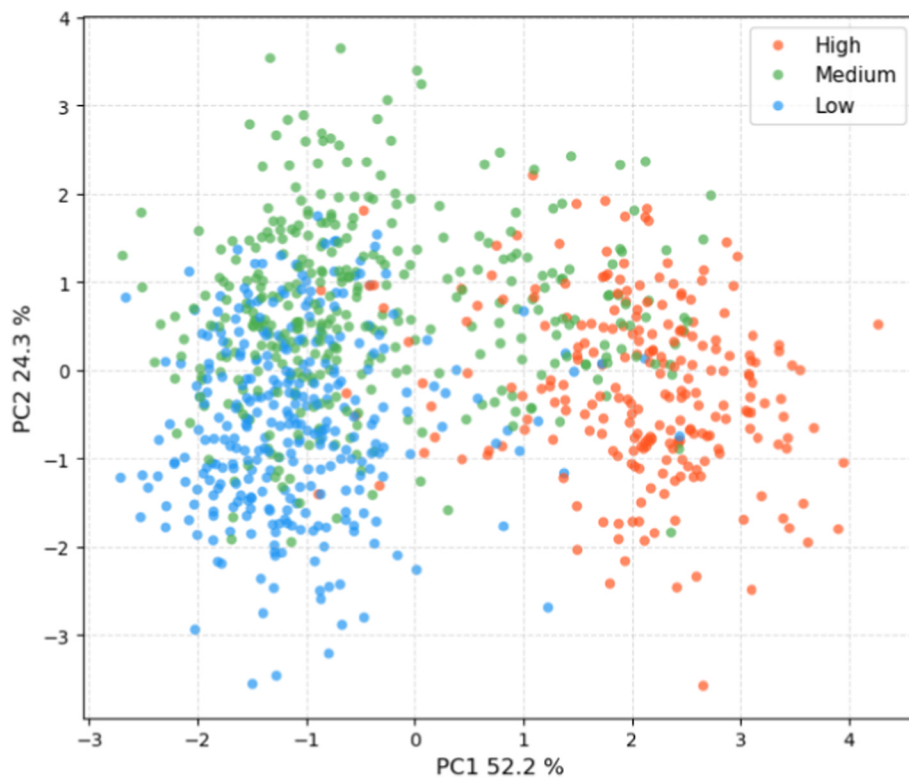


Fig. 7. PCA analysis of the spectra showing differences based on the internal cluster pigmentation.

with O–H and C–H combination bands related to water and organic compound content. And PC4 (7 %) and PC5 (4 %) captured additional structured variance around 1350–1450 nm (Nicolai et al., 2007). Together, these results indicate that whilst the between-group spectral differences are subtle, they are systematic and distributed across the NIR range, providing a sufficient spectral basis for the machine learning models to discriminate between pigmentation levels.

To better characterise the underlying spectral differences among samples according to their composition or grouping, the results of an

unsupervised analysis, using PCA are shown in Fig. 7. The ability of PCA to reveal spectral separation concordant with internal anthocyanin clustering is consistent with previous studies demonstrating that unsupervised chemometric analysis of NIR data can effectively discriminate samples based on compositional differences without a priori class information (Duntelman, 1989).

The first two principal components (PC1 and PC2) together explained 76.5% of the total spectral variance (52.2% and 24.3%, respectively), confirming that the dominant sources of variation within

the dataset were adequately captured in a reduced two-dimensional space. Fig. 7 illustrates that the PCA score plot separated the spectra according to the defined clusters, with the most evident discrimination occurring between samples with high and low anthocyanin content, whereas the distinction between low- and medium-pigmented oranges was less pronounced. The partial overlap between low and medium samples suggests that while these groups share similar anthocyanin levels, their secondary separation is likely driven by variations in co-occurring metabolites, such as soluble sugars, water distribution, and other phenolic compounds which are known to influence the near-infrared spectral profile of citrus fruits (Elhady et al., 2024).

These findings confirm that the spectral data contain sufficient discriminatory information to distinguish oranges according to their internal coloration profile, providing a strong unsupervised basis for the subsequent supervised classification models.

### 3.3. Estimation of fruit quality attributes using non-destructive prediction models

Internal quality estimation is highly beneficial for packing houses and essential for consumers. Table 4 shows the prediction metrics for TSS, TA, MI, and total anthocyanin content across the three tested models. All models exhibited similar overall performances. The top predictions were achieved for TSS, with the MLP model producing the best results, with an  $R^2$  value of 0.724. The PLS and SVM models followed closely behind, with  $R^2$  values of 0.711 and 0.691, respectively. For anthocyanin prediction, the MLP model also outperformed the others, achieving an  $R^2$  value of 0.613. This success was due to the ability of neural networks to identify non-linear relationships in the data, whereas the PLS and SVM models showed  $R^2$  values of 0.591 and 0.568, respectively. On the other hand, predictions for TA and MI were poor, with the highest  $R^2$  values being only 0.323 and 0.366, for MLP and PLS models respectively.

### 3.4. Estimation of internal pigmentation using non-destructive prediction models

The spectral differences observed in the previous exploratory analysis indicated that internal pigmentation was associated with distinguishable spectral patterns; therefore, classification and model training were conducted to assess the feasibility of predicting internal pigmentation from the hyperspectral data.

When predicting the three distinct pigmentation groups, the non-linear models (SVM and MLP) consistently outperformed the linear PLS model, as shown in Table 5. During training, MLP and SVM reached average accuracies of 0.92 and 0.90, respectively, demonstrating their

strong capacity to learn complex spectral patterns, while PLS achieved an average training accuracy of 0.81. In validation, the SVM and MLP models maintained an overall average accuracy of 0.77, compared to the 0.76 accuracy achieved by the PLS model.

However, a detailed examination of the confusion matrices (Table 6) and performance metrics reveals distinct differences in how the SVM and MLP models classify the various groups. The SVM model proved superior at identifying highly pigmented samples, achieving the highest validation accuracy (0.77), precision (0.70), and recall (0.52) for the “High” pigmentation class. This allowed for more precise differentiation between deeply red and less pigmented blood oranges. Conversely, the MLP model was more effective at classifying oranges from the low and medium pigmentation groups. In validation, MLP achieved the highest precision (0.76) and recall (0.88) for the “Low” class, as well as the highest precision (0.70) for the “Medium” class. Despite this strength, the MLP model struggled more with the “High” class, frequently misclassifying high-pigmented oranges as medium-pigmented.

Predicting red and non-red blood oranges based on two distinct pigmentation groups performed better than attempting to separate them into three distinct groups. By merging the ambiguous medium class, the algorithms established a much clearer decision boundary. As shown in Table 7, during training the MLP and SVM achieving high average accuracies of 0.92 and 0.91, respectively, while the PLS model reached 0.83. This strong performance translated into highly consistent results across all models during validation. Both the PLS and MLP models achieved an average validation accuracy of 0.79, while the SVM followed closely with 0.78.

However, the confusion matrices (Table 8) and detailed metrics highlight subtle operational differences between the models. The MLP model was slightly more balanced overall and was effective in predicting samples with lower pigmentation values, maintaining a solid validation precision of 0.67 for the “Low” class. On the contrary, the SVM excelled in correctly classifying samples with higher pigmentation levels, as in the three-pigmentation classification case, achieving an excellent validation precision of 0.96 for the “High” group, meaning that when the SVM labeled an orange as highly pigmented, it was almost always correct. Interestingly, while the PLS model is the simplest, it performed remarkably well in this binary task, matching the MLP's overall validation accuracy and achieving the highest validation recall (0.94) for the “Low” class.

## 4. Discussion

The observed variations in anthocyanin content, total soluble solids (TSS), and titratable acidity (TA) across the studied seasons highlighted the complex physiological responses of citrus to environmental

**Table 4**

Training and validation performance metrics of PLS, SVM, and MLP regression models for non-destructive prediction of fruit quality parameters. <sup>Y</sup>

| Parameter    | Train |                |       |       | Validation     |       |        |
|--------------|-------|----------------|-------|-------|----------------|-------|--------|
|              | Model | R <sup>2</sup> | MAE   | RMSE  | R <sup>2</sup> | MAE   | RMSE   |
| TSS          | PLS   | 0.763          | 0.384 | 0.492 | 0.711          | 0.439 | 0.559  |
|              | SVM   | 0.802          | 0.342 | 0.448 | 0.691          | 0.457 | 0.586  |
|              | MLP   | 0.841          | 0.315 | 0.412 | 0.724          | 0.435 | 0.554  |
| TA           | PLS   | 0.385          | 0.112 | 0.141 | 0.297          | 0.131 | 0.164  |
|              | SVM   | 0.456          | 0.098 | 0.122 | 0.309          | 0.123 | 0.158  |
|              | MLP   | 0.512          | 0.091 | 0.114 | 0.323          | 0.119 | 0.155  |
| MI           | PLS   | 0.442          | 0.854 | 1.085 | 0.366          | 0.949 | 1.207  |
|              | SVM   | 0.487          | 0.793 | 1.032 | 0.358          | 0.942 | 1.215  |
|              | MLP   | 0.529          | 0.751 | 0.984 | 0.341          | 0.957 | 1.235  |
| Anthocyanins | PLS   | 0.654          | 7.341 | 9.215 | 0.591          | 8.058 | 10.337 |
|              | SVM   | 0.712          | 6.582 | 8.453 | 0.568          | 8.398 | 10.619 |
|              | MLP   | 0.785          | 5.924 | 7.641 | 0.613          | 7.854 | 9.963  |

<sup>Y</sup> Abbreviations: TSS (total soluble solids); TA (titratable acidity); MI (maturity index);  $R^2$  (R-squared/coefficient of determination); MAE (mean absolute error); and RMSE (root mean square error).

**Table 5**

Class-wise and overall performance metrics for the three-class pigmentation prediction models across training and validation sets.

| Model | Pigmentation | Training  |        |          | Validation |        |          |
|-------|--------------|-----------|--------|----------|------------|--------|----------|
|       |              | Precision | Recall | Accuracy | Precision  | Recall | Accuracy |
| PLS   | Low          | 0.79      | 0.8    | 0.86     | 0.74       | 0.86   | 0.85     |
|       | Medium       | 0.7       | 0.64   | 0.72     | 0.62       | 0.5    | 0.65     |
|       | High         | 0.61      | 0.7    | 0.85     | 0.47       | 0.54   | 0.78     |
|       | Average      | 0.7       | 0.72   | 0.81     | 0.61       | 0.63   | 0.76     |
| SVM   | Low          | 0.89      | 0.9    | 0.93     | 0.75       | 0.87   | 0.85     |
|       | Medium       | 0.78      | 0.82   | 0.87     | 0.52       | 0.56   | 0.69     |
|       | High         | 0.91      | 0.84   | 0.92     | 0.7        | 0.52   | 0.77     |
|       | Average      | 0.86      | 0.86   | 0.9      | 0.66       | 0.65   | 0.77     |
| MLP   | Low          | 0.91      | 0.93   | 0.94     | 0.76       | 0.88   | 0.86     |
|       | Medium       | 0.91      | 0.84   | 0.89     | 0.7        | 0.52   | 0.67     |
|       | High         | 0.77      | 0.88   | 0.93     | 0.42       | 0.57   | 0.79     |
|       | Average      | 0.86      | 0.89   | 0.92     | 0.63       | 0.66   | 0.77     |

**Table 6**

Training and validation confusion matrices for the three-class pigmentation prediction models across training and validation sets.

|            |           |  | Models |        |      |      |        |      |      |        |      |
|------------|-----------|--|--------|--------|------|------|--------|------|------|--------|------|
|            |           |  | PLS    |        |      | SVM  |        |      | MLP  |        |      |
|            |           |  | Real   |        |      | Real |        |      | Real |        |      |
| Training   | Predicted |  | Low    | Medium | High | Low  | Medium | High | Low  | Medium | High |
|            | Low       |  | 228    | 54     | 6    | 257  | 25     | 7    | 268  | 24     | 3    |
|            | Medium    |  | 56     | 230    | 44   | 27   | 219    | 35   | 19   | 322    | 14   |
|            | High      |  | 0      | 74     | 117  | 0    | 22     | 217  | 0    | 36     | 123  |
| Validation | Predicted |  | Low    | Medium | High | Low  | Medium | High | Low  | Medium | High |
|            | Low       |  | 82     | 26     | 3    | 83   | 20     | 8    | 84   | 24     | 3    |
|            | Medium    |  | 13     | 60     | 23   | 12   | 50     | 34   | 12   | 67     | 17   |
|            | High      |  | 0      | 34     | 30   | 0    | 19     | 45   | 0    | 37     | 27   |

**Table 7**

Class-wise and overall performance metrics for the two-class pigmentation prediction models across training and validation sets.

| Model | Training     |           |        |          | Validation |        |          |
|-------|--------------|-----------|--------|----------|------------|--------|----------|
|       | Pigmentation | Precision | Recall | Accuracy | Precision  | Recall | Accuracy |
| PLS   | Low          | 0.74      | 0.84   | 0.83     | 0.65       | 0.94   | 0.79     |
|       | High         | 0.9       | 0.83   | 0.83     | 0.95       | 0.7    | 0.79     |
|       | Average      | 0.82      | 0.84   | 0.83     | 0.8        | 0.82   | 0.79     |
| SVM   | Low          | 0.84      | 0.93   | 0.91     | 0.62       | 0.95   | 0.78     |
|       | High         | 0.96      | 0.9    | 0.91     | 0.96       | 0.69   | 0.78     |
|       | Average      | 0.9       | 0.92   | 0.91     | 0.79       | 0.82   | 0.78     |
| MLP   | Low          | 0.88      | 0.94   | 0.92     | 0.67       | 0.9    | 0.79     |
|       | High         | 0.96      | 0.91   | 0.92     | 0.91       | 0.71   | 0.79     |
|       | Average      | 0.92      | 0.93   | 0.92     | 0.79       | 0.8    | 0.79     |

**Table 8**

Training and validation confusion matrices for the two-class pigmentation prediction models across training and validation sets.

|            |           |  | Models |      |      |      |      |      |
|------------|-----------|--|--------|------|------|------|------|------|
|            |           |  | PLS    |      | SVM  |      | MLP  |      |
|            |           |  | Real   |      | Real |      | Real |      |
| Training   | Predicted |  | Low    | High | Low  | High | Low  | High |
|            | Low       |  | 250    | 88   | 265  | 52   | 303  | 43   |
|            | High      |  | 46     | 425  | 19   | 473  | 19   | 444  |
| Validation | Predicted |  | Low    | High | Low  | High | Low  | High |
|            | Low       |  | 93     | 51   | 90   | 54   | 97   | 47   |
|            | High      |  | 6      | 121  | 5    | 122  | 11   | 116  |

conditions. While [Tarancón et al. \(2022\)](#) reported a positive correlation between pulp pigmentation and TSS, the data from the present study demonstrated that excessive rainfall could decouple the accumulation of sugars and anthocyanins. This aligned with the findings of [Sadras and Moran \(2012\)](#), who noted that sugar accumulation and secondary

metabolite biosynthesis were regulated by independent physiological pathways. Consequently, anthocyanin levels remained relatively resilient to water-induced dilution, whereas TSS was significantly impacted.

Furthermore, the results confirmed that temperature was a primary driver of anthocyanin accumulation, consistent with previous studies

indicating that cold stress activated the genetic pathways responsible for pigmentation in blood oranges (Caruso et al., 2016; Chen et al., 2023; Crifò et al., 2011; Rapisarda et al., 2001). Specifically, a strong correlation ( $r = 0.73$ ) was observed between total anthocyanin content and the number of chilling hours (below 4 °C) accumulated prior to harvest. This reflected the mechanism originally elucidated by Butelli et al. (2012), wherein cold exposure triggered the Ruby transcriptional regulator. Even when modern biostimulants were applied, low temperatures remained the essential catalyst for pigment development in ‘Sanguinelli’ oranges (Garrido-Auñón, Padilla-González, et al., 2025, Garrido-Auñón, Padilla-González, et al., 2025). Consequently, variations in the timing and duration of chilling hours between seasons directly explained the inter-annual heterogeneity observed in the dataset.

Despite this heterogeneity, the predictive models developed in this study achieved reliable results, but their performance was dependent on the validation strategy. The models achieved accuracy comparable to previous studies, such as Pires et al. (2022), who reported  $R^2$  values around 0.80 for TSS and TA in Ortanique citrus using data from multiple seasons. However, there is a strong influence of climatic variability. Declines were reported by Guthrie et al. (2005) for mandarins ( $R^2$  dropping to 0.35) and by Zude et al. (2008), who highlighted the necessity of multi-season recalibration to account for precipitation effects. The results were also consistent with Serna-Escolano, Giménez, Zapata, et al. (2024), who observed that external calibration in lemons reduced the  $R^2$  for TA to 0.54. This underscored that, even with a 3 year dataset, there is a critical need for broad, multi-season datasets to build robust agricultural models.

Crucially, the ability of near-infrared (NIR) spectroscopy to predict internal parameters such as TSS and anthocyanins raised important questions regarding light penetration. Because NIR radiation has limited penetration depth in fruit, as proved in Persimmon fruit by Rodríguez-Ortega et al. (2023) or in avocado Wedding et al. (2024), the spectral signal primarily originated from the flavedo and albedo layers. However, the ‘Sanguinelli’ blood orange variety possessed a relatively thin rind compared to other citrus cultivars. The rind thickness of these fruits often fell within or near the upper limit of NIR diffuse reflectance penetration, allowing the light to interact more directly with the outer margins of the internal pulp (Borba et al., 2020). This physical characteristic facilitated the extraction of internal information that would be significantly more challenging, or even impossible, to acquire in thick-skinned varieties like navel oranges, where the rind acts as a severe barrier to light penetration (Aline et al., 2023; Song et al., 2019). Furthermore, the physiological triggers for pigmentation in blood oranges (such as cold stress) simultaneously affect the biochemical composition of the rind altering its density, moisture, and secondary metabolite profile (Dong et al., 2022; Song et al., 2019). The models likely exploited both the partial direct penetration into the pulp and these co-varying superficial features as proxies for internal quality. While this combination of direct and indirect relationships enabled effective non-destructive sorting, it also explained the models' vulnerability to seasonal climatic shifts, which could alter rind characteristics independently of internal pulp composition.

Future studies should aim to incorporate data across more diverse environmental conditions and harvesting seasons to improve model transferability. Nonetheless, the current work demonstrated the practical viability of sorting blood oranges based on internal pigmentation, providing packing houses with a valuable tool to non-destructively separate highly pigmented, high-value fruits from lighter ones.

## 5. Conclusions

This study demonstrated that the internal quality and pigmentation of ‘Sanguinelli’ blood oranges are highly dependent on seasonal climate, with early chilling hours critically driving anthocyanin biosynthesis. While near-infrared evaluation effectively predicted TSS ( $R^2 = 0.72$ ) and

total anthocyanins ( $R^2 = 0.61$ ) using a MLP model, predictions for TA and MI were less satisfactory. A primary limitation identified is the models' vulnerability to seasonal variations. Because the spectral signal partially relies on co-varying superficial rind features as proxies for internal quality, independent environmental factors (such as heavy rainfall) can alter the rind and impact model accuracy.

Despite this limitation, the MLP model achieved 79% accuracy in classifying oranges by internal pigmentation. From an industrial perspective, this confirms the practical viability of non-destructive spectral sorting in commercial packing houses, enabling the automated segregation of high-value, deep-red fruits to meet premium market demands. To guarantee reliable performance in future seasons under variable climatic conditions, subsequent research must integrate broad, multi-season datasets to enhance model robustness and commercial transferability.

## CRedit authorship contribution statement

**Salvador Castillo-Girones:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **María E. García-Pastor:** Validation, Methodology, Investigation. **María Serrano:** Writing – review & editing, Visualization, Validation, Resources, Investigation, Funding acquisition. **Daniel Valero:** Writing – review & editing, Visualization, Validation, Resources, Project administration, Investigation, Funding acquisition. **Sergio Cubero:** Validation, Methodology, Investigation, Formal analysis. **Salvador Castillo:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation, Conceptualization.

## Funding

The authors declare financial support was received for the research and/or publication of this article. This research was funded by Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital of the Generalitat Valenciana, Spain, through the project PROMETEO/2021/089 “Innovative and eco-friendly pre- and postharvest strategies with natural compounds to improve quality of fruits”.

## 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.

## Acknowledgments

The authors thank Direcció General de Ciència i Investigació of the Generalitat Valenciana, Spain, for the financial support through the Prometeo Program (PROMETEO/2021/089) and the company Las Moreras Fruits & Veggies for permission to use their plots and for their technical support. Finally, the authors thank BioRender's free trial (Toronto, ON, Canada) for the images used in the Graphical Abstract.

## Data availability

Data will be made available on request.

## References

- AEMET - Estación meteorológica Rojales, El Molino. (n.d.). Retrieved February 8, 2026, from <https://x-y.es/aemet/est-7261X-rojales-el-molino#indiceP1>.
- Aline, U., Bhattacharya, T., Faqeerzada, M. A., Kim, M. S., Baek, I., & Cho, B. K. (2023). Advancement of non-destructive spectral measurements for the quality of major tropical fruits and vegetables: A review. *Frontiers in Plant Science*, 14, Article 1240361. <https://doi.org/10.3389/FPLS.2023.1240361>
- Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, 174, 33–44. <https://doi.org/10.1016/J.CHEMOLAB.2017.12.004>

- Borba, K. R., Spricigo, P. C., Aykas, D. P., Mitsuyuki, M. C., Colnago, L. A., & Ferreira, M. D. (2020). Non-invasive quantification of vitamin C, citric acid, and sugar in 'Valência' oranges using infrared spectroscopies. *Journal of Food Science and Technology* 2020, 58(2), 731–738. <https://doi.org/10.1007/S13197-020-04589-X>, 58(2).
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., & Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell*, 24(3), 1242–1255. <https://doi.org/10.1105/tpc.111.095232>
- Cai, L., Li, J., Zhang, H., Zhang, Y., Zhan, J., & Hao, H. (2025). Determination of the SSC in oranges using Vis-NIR full transmittance hyperspectral imaging and spectral visual coding: A practical solution to the scattering problem of inhomogeneous mixtures. *Food Chemistry*, 474, Article 143239. <https://doi.org/10.1016/j.foodchem.2025.143239>, 2025.
- Caruso, M., Ferlito, F., Licciardello, C., Allegra, M., Strano, M. C., Di Silvestro, S., Russo, M. P., Pietro Paolo, D., Caruso, P., Las Casas, G., Stagno, F., Torrisi, B., Rocuzzo, G., Reforgiato Recupero, G., & Russo, G. (2016). Pomological diversity of the Italian blood orange germplasm. *Scientia Horticulturae*, 213, 331–339. <https://doi.org/10.1016/j.scienta.2016.10.044>
- Cavaco, A. M., Pires, R., Antunes, M. D., Panagopoulos, T., Brázio, A., Afonso, A. M., Silva, L., Lucas, M. R., Cadeiras, B., Cruz, S. P., & Guerra, R. (2018). Validation of short wave near infrared calibration models for the quality and ripening of 'Newhall' orange on tree across years and orchards. *Postharvest Biology and Technology*, 141, 86–97. <https://doi.org/10.1016/j.postharvbio.2018.03.013>
- Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18(2), 72–83. <https://doi.org/10.1016/j.tifs.2006.09.003>
- Chen, J., Zhang, Y., Liu, F., Chen, J., Wang, W., Wu, D., Ye, X., Liu, D., & Cheng, H. (2023). The potential of different ripeness of blood oranges (*Citrus sinensis* L. Osbeck) for sale in advance after low-temperature storage: Anthocyanin enhancements, volatile compounds, and taste attributes. *Food Chemistry*, 417, Article 135934. <https://doi.org/10.1016/j.foodchem.2023.135934>
- Crifo, T., Puglisi, I., Petrone, G., Recupero, G. R., & Lo Piero, A. R. (2011). Expression analysis in response to low temperature stress in blood oranges: Implication of the flavonoid biosynthetic pathway. *Gene*, 476(1–2), 1–9. <https://doi.org/10.1016/j.gene.2011.02.005>
- Custodio-Mendoza, J. A., Aktas, H., Zalewska, M., Wyrwiz, J., & Kurek, M. A. (2024). A review of quantitative and topical analysis of anthocyanins in food. *Molecules*, 29(8), 1735. <https://doi.org/10.3390/molecules29081735>
- Dong, Y., Shan, Y., Li, P., Jiang, L., & Liu, X. (2022). Nondestructive characterization of citrus fruit by near-infrared diffuse reflectance spectroscopy (NIRDRS) with principal component analysis (PCA) and fisher linear discriminant analysis (FLDA). *Analytical Letters*, 55(16), 2554–2563. <https://doi.org/10.1080/00032719.2022.2063306>
- Dunteman, G. H. (1989). *Principal components analysis*. <https://doi.org/10.4135/9781412985475>
- Elhady, S. S., Youssef, F. S., Lashkar, M. O., Hamdan, D. I., Ashour, M. L., Zengin, G., & Gamal El-Din, M. I. (2024). Chemometric discrimination of eight citrus plants utilizing chromatographic and spectroscopic techniques and insights into their biological potentials. *Current Research in Food Science*, 9, Article 100803. <https://doi.org/10.1016/j.crf.2024.100803>
- Eurostat. Orange, lemon and small citrus fruit trees-area by age and density classes (area in ha). <https://ec.europa.eu/eurostat>
- Fearn, T., Riccioli, C., Garrido-Varo, A., & Guerrero-Ginel, J. E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, 96(1), 22–26. <https://doi.org/10.1016/j.chemolab.2008.11.006>
- Garrido-Auñón, F., Padilla-González, P. A., Serrano, M., Valero, D., & Agulló, V. (2025). Melatonin boosts the phytochemical profile of blood oranges, enhancing (poly) phenol and endogenous melatonin content, through pre- and postharvest treatments. *Journal of Pineal Research*, 77(5), Article e70078. <https://doi.org/10.1111/jpi.70078>
- Garrido-Auñón, F., Puente-Moreno, J., García-Pastor, M. E., Zapata, P. J., Martínez-Romero, D., Serrano, M., & Valero, D. (2025). Brassinosteroid preharvest treatments as a useful tool to increase crop yield and red colour in blood orange fruits. *Frontiers in Plant Science*, 16, Article 1654517. <https://doi.org/10.3389/fpls.2025.1654517>
- Giménez-Sanchis, A., Pons-Gómez, A., Farina, V., & Besada, C. (2022b). Effect of pulp pigmentation intensity on consumer acceptance of new blood mandarins: A cross-cultural study in Spain and Italy. *Agronomy*, 12(12), 3058. <https://doi.org/10.3390/AGRONOMY12123058>. Page 3058, 12.
- Giménez-Sanchis, A., Zhong, K., Pintor, A., Farina, V., & Besada, C. (2022a). Understanding blood versus blond orange consumption: A cross-cultural study in four countries. *Foods*, 11(17), 2686. <https://doi.org/10.3390/FOODS11172686>
- Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers* 2022, 2(1), 1–21. <https://doi.org/10.1038/s43586-022-00184-w>, 2(1).
- Grosso, G., Galvano, F., Mistretta, A., Marventano, S., Nolfo, F., Calabrese, G., Buscemi, S., Drago, F., Veronesi, U., & Scuderi, A. (2013). Red orange: Experimental models and epidemiological evidence of its benefits on human health. *Oxidative Medicine and Cellular Longevity*, 2013(1), Article 157240. <https://doi.org/10.1155/2013/157240>
- Guthrie, J. A., Reid, D. J., & Walsh, K. B. (2005). Assessment of internal quality attributes of mandarin fruit. 2. NIR calibration model robustness. *Australian Journal of Agricultural Research*, 56(4), 417–426. <https://doi.org/10.1071/AR04299>
- Habibi, F., Ramezani, A., Guillén, F., Castillo, S., Serrano, M., & Valero, D. (2020). Changes in bioactive compounds, antioxidant activity, and nutritional quality of blood orange cultivars at different storage temperatures. *Antioxidants*, 9(10), 1016. <https://doi.org/10.3390/ANTIOX9101016>, 1016, 9.
- Lo Piero, A. R. (2015). The state of the art in biosynthesis of anthocyanins and its regulation in pigmented sweet oranges [*Citrus sinensis* L. Osbeck]. *Journal of Agricultural and Food Chemistry*, 63(16), 4031–4041. <https://doi.org/10.1021/acs.jafc.5B01123>
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python : A guide for data scientists, 378 <https://www.oreilly.com/library/view/introduction-to-machine-learning/9781449369880/>.
- Madhulatha, T. S. (2012). An overview on clustering methods. *IOSR Journal of Engineering*, 2(4), 719–725. <https://doi.org/10.9790/3021-0204719725>
- Moomkesh, S., Mireei, S. A., Sadeghi, M., & Nazeri, M. (2017). Early detection of freezing damage in sweet lemons using Vis/SWNIR spectroscopy. *Biosystems Engineering*, 164, 157–170. <https://doi.org/10.1016/j.biosystemseng.2017.10.009>
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>
- Pan, L., Zhang, Q., Zhang, W., Sun, Y., Hu, P., & Tu, K. (2016). Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chemistry*, 192, 134–141. <https://doi.org/10.1016/j.foodchem.2015.06.106>
- Pasquini, C. (2003). Near Infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Pérez-Pérez, J. G., Cámara-Campos, E., Giménez-Sanchis, A., Badal, E., Martínez-Gimeno, M. A., Tasa, M., Bonet, L., & Besada, C. (2023). Efectos del riego deficitario controlado sobre la coloración del fruto en naranja sanguina. IX Congreso Ibérico y XVII Congreso Nacional de Ciencias Hortícolas, 93, 366–369. [https://redivia.gva.es/bitstream/handle/20.500.11939/8851/2023\\_Pérez-Pérez\\_Efectos.pdf?sequence=1&isAllowed=y](https://redivia.gva.es/bitstream/handle/20.500.11939/8851/2023_Pérez-Pérez_Efectos.pdf?sequence=1&isAllowed=y).
- Pio Ávila, B., Cardozo, L. O., Alves, G. D., Pereira, A. M., Gularte, M. A., & de Oliveira, R. P. (2022). Targeted chemical and sensory profiling to guide consumption of blood orange. *Journal of Culinary Science & Technology*, 20(4), 350–365. <https://doi.org/10.1080/15428052.2020.1843581>
- Pires, R., Guerra, R., Cruz, S. P., Antunes, M. D., Brázio, A., Afonso, A. M., Daniel, M., Panagopoulos, T., Gonçalves, L., & Cavaco, A. M. (2022). Ripening assessment of 'Ortanique' (*Citrus reticulata* Blanco x *Citrus sinensis* (L) Osbeck) on tree by SW-NIR reflectance spectroscopy-based calibration models. *Postharvest Biology and Technology*, 183, Article 111750. <https://doi.org/10.1016/j.postharvbio.2021.111750>
- Rapisarda, P., Bellomo, S. E., & Intelisano, S. (2001). Storage temperature effects on blood orange fruit quality. *Journal of Agricultural and Food Chemistry*, 49(7), 3230–3235. <https://doi.org/10.1021/JF010032L>
- Riccioli, C., Pérez-Marín, D., & Garrido-Varo, A. (2021). Optimizing spatial data reduction in hyperspectral imaging for the prediction of quality parameters in intact oranges. *Postharvest Biology and Technology*, 176, Article 111504. <https://doi.org/10.1016/j.postharvbio.2021.111504>
- Rodríguez-Ortega, A., Aleixos, N., Blasco, J., Albert, F., & Munera, S. (2023). Study of light penetration depth of a Vis-NIR hyperspectral imaging system for the assessment of fruit quality. A case study in persimmon fruit. *Journal of Food Engineering*, 358, Article 111673. <https://doi.org/10.1016/j.jfoodeng.2023.111673>
- Sadras, V. O., & Moran, M. A. (2012). Elevated temperature decouples anthocyanins and sugars in berries of Shiraz and Cabernet Franc. *Australian Journal of Grape and Wine Research*, 18(2), 115–122. <https://doi.org/10.1111/j.1755-0238.2012.00180.x>
- Serna-Escolano, V., Giménez, M. J., Serrano, M., Valero, D., García-Pastor, M. E., Dobón-Suarez, A., Gutiérrez-Pozo, M., Giménez-Berenguer, M., & Zapata, P. J. (2024). Fruit position on tree canopy affects fruit quality traits in 'Sanguinelli' blood oranges. *Horticulturae*, 10, 949. <https://doi.org/10.3390/horticulturae10090949>
- Serna-Escolano, V., Giménez, M. J., Zapata, P. J., Cubero, S., Blasco, J., & Munera, S. (2024). Non-destructive assessment of "Fino" lemon quality through ripening using NIRS and chemometric analysis. *Postharvest Biology and Technology*, 212, Article 112870. <https://doi.org/10.1016/j.postharvbio.2024.112870>
- Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning Applications-A vision. *Global Transitions Proceedings*, 2(1), 24–28. <https://doi.org/10.1016/j.gltp.2021.01.004>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Song, J., Li, G., & Yang, X. (2019). Optimizing genetic algorithm-partial least squares model of soluble solids content in Fukumoto navel orange based on visible-near-infrared transmittance spectroscopy using discrete wavelet transform. *Journal of the Science of Food and Agriculture*, 99(11), 4898–4903. <https://doi.org/10.1002/JSCA.9717>
- 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>
- Tarancon, P., Cebrián, B., Fernández-Serrano, P., & Besada, C. (2022). Relation between rind pigmentation and internal quality of blood orange 'Sanguinelli': Physicochemical and Sensory studies. *Horticulturae*, 8(5), 448. <https://doi.org/10.3390/HORTICULTURAE8050448>
- Tian, S., Wang, S., & Xu, H. (2022). Early detection of freezing damage in oranges by online Vis/NIR transmission coupled with diameter correction method and deep 1D-CNN. *Computers and Electronics in Agriculture*, 193, Article 106638. <https://doi.org/10.1016/j.compag.2021.106638>
- Vranić, M., Bošnjak, K., Rukavina, I., Glavanović, S., Pintić Pukec, N., Babić, A., & Vranić, I. (2020). Prediction of forage chemical composition by NIR spectroscopy. *Journal of Central European Agriculture*, 21(3), 554–568. <https://doi.org/10.5513/JCEA01/21.3.2839>

- Wedding, B. B., Wright, C., Grauf, S., White, R. D., Wedding, B. B., Wright, C., Grauf, S., & White, R. D. (2024). Wavelength variation of the depth of penetration of near infrared radiation in “Hass” avocado fruit. *Technology in Horticulture*, 4(1). <https://doi.org/10.48130/THORT-0024-0005>, 1:E008. 0–0.
- Wu, D., & Sun, D. W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. *Innovative Food Science & Emerging Technologies*, 19, 1–14. <https://doi.org/10.1016/J.IFSET.2013.04.014>
- Zude, M., Pflanz, M., Kapielian, C., & Aivazian, B. L. (2008). NIRS as a tool for precision horticulture in the citrus industry. *Biosystems Engineering*, 99(3), 455–459. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2007.10.016>



**DSPACE**

<https://dspace.org/>

## **Non-destructive quality estimation and internal pigmentation discrimination of 'Sanguinelli' blood oranges using hyperspectral imaging**

**Castillo-Gironés, Salvador; García-Pastor, María E.; Serrano, María; Valero, Daniel; Cubero, Sergio ...**

**2026-11**

Elsevier

<https://doi.org/10.1016/j.foodcont.2026.112324>

Castillo-Girones S, García-Pastor ME, Serrano M, et al., (2026) Non-destructive quality estimation and internal pigmentation discrimination of 'Sanguinelli' blood oranges using hyperspectral imaging. Food Control, Volume 189, November 2026, Article number 112324