Volume 2, 2026, Pages 150-161 *Journal of Smart and Sustainable Farming*
| Regular Article | Open Access

Intelligent Post Harvest Quality Prediction of Fruits Using Non-Destructive Spectral Analysis

Lise Roy Porter
Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, South Africa.
lise.porter@up.ac.za

Mekdes Solomon Cherinet
Jimma University, Jimma, Ethiopia.
mekdes.solomon@ju.edu.et

Article Info
Journal of Smart and Sustainable Farming
https://www.ansispublications.com/jssf/jssf.html

Received 04 April 2026
Revised from 06 May 2026
Accepted 30 June 2026
Available online 18 July 2026
Published by Ansis Publications.

© The Author(s), 2026.
https://doi.org/10.64026/JSSF/2026013

Corresponding author(s):
Mekdes Solomon Cherinet, Jimma University, Jimma, Ethiopia.
Email: mekdes.solomon@ju.edu.et

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/ licenses/by/4.0/).

Abstract – Post-harvest losses of fruit are mainly caused by the absence of external quality as an indicator of internal quality (ripeness, eating quality and remaining shelf life) which are parameters that cannot be assessed from external inspection, and the conventional methods used to assess these parameters (soluble solids content, SSC and firmness) are destructive, slow and require only small portions of fruit. An intelligent framework has been developed to predict the internal quality and storage stage of mango using the diffuse reflectance spectra (DRS) (600–1700 nm) of the fruit in the visible–near-infrared (Vis–NIR) region, obtained through diffuse reflectance spectroscopy (DRS), and hyperspectral images. The framework integrates absorbance conversion, scatter correction (standard normal variate, SNV), Savitzky–Golay (SG) derivatives, partial least squares regression (PLSR), iterative wavelength elimination and machine-learning regressors with the extension to point predictions to pixel-wise quality maps and first-order shelf-life kinetics. Within the set of 240 fruits stored for 0–14 days (160 were used for calibration and 80 were used for independent prediction), the PLSR model with SNV + SG first-derivative spectra predicted SSC with R2P = 0.91, RMSEP = 0.93 °Brix and RPD = 3.3, while firmness was predicted with R2P = 0.92, RMSEP = 4.6 N and RPD = 3.6. The accuracy of the firmness remained similar when 36 of 221 wavelengths were selected, and this accuracy was comparable to that for a support vector regressor, and inferior to that for a multilayer perceptron, with the available sample size. The three maturity stages were classified correctly with 75 % accuracy and the maturity stage at which firmness dropped to 15 N was predicted with 8.2 d against 8.1 d measured. The method allows for fast grading, control of ripening room conditions and logistics decisions based on the shelf life without harming the fruit.

Keywords – Non-Destructive Testing, Near-Infrared Spectroscopy, Hyperspectral Imaging, Post-Harvest Quality, Partial Least Squares, Mango, Shelf Life.

I. INTRODUCTION

Fruit is highly perishable and a significant proportion of fruit produced is lost or downgraded after harvest as ripening, softening and decay occur after harvesting [1]. The problem is compounded by the fact that the sugar content, firmness and shelf life of two fruits of the same size and skin colour can vary greatly, depending on stage of maturity at harvest, orchard growing conditions and storage history. Such decisions as market, length of cold storage, timing of ripening treatment, and fruit selling price often are therefore based on the appearance of the fruit from outside, experience, and small destructive samples, which are not representative of the lot.

Non-destructive spectral analysis is a solution to this information gap. Light in the visible and near-infrared range interacts with fruit tissue by absorption due to pigments, absorption due to O–H, C–H and N–H bonds in water, sugars, acids and cell-wall polymers, and scattering by the cellular microstructure. The diffusely reflected spectrum is thus a "fingerprint" of the chemical composition and texture, then a calibration model can be developed to convert the spectrum

to a value of an internal quality attribute. There has been extensive review of the use of near-infrared spectroscopy for the non-destructive assessment of fruit and vegetable quality [2] and application of the technique to specific commodities such as citrus demonstrates the potential for external quality assessment as well as internal quality [4]. Point spectroscopy in the visible–NIR has been commercially applied for about 30 years to sort fruit on the basis of internal quality [3] and the geometry of the measurement (reflectance, interactance or transmission) has a significant influence on the depth of penetration of the light into the fruit flesh [5]. Hyperspectral imaging provides spatial information on the spectrum and is a new analytical tool for food quality control since it can map food properties over the surface of the fruit [6].

There are three practical issues that remain, however. One, a large proportion of the spectra consists of broad, overlapping and highly correlated bands, and scatter effects are not strictly related to chemistry, which means that the pre-processing of spectra and the use of certain wavelengths have an impact on the accuracy and robustness of the calibration [7]. Second, the model class: linear latent-variable models are common, but other models are increasingly being suggested (e.g., machine-learning or deep-learning models), and their performance in smaller data sets, more commonly encountered in post-harvest research, is not always evident. Third, most studies end up predicting just one attribute at just one point; a grower or packhouse, however, needs to make integrated decisions – which fruit to grade out, what stage the fruit has reached, how it is distributed within the fruit and, most importantly, how many days are left before it becomes unmarketable.

This paper thus proposes and tests an intelligent, end-to-end solution for the post-harvest quality prediction of Vis-NIR spectra. The goals are to (1) quantify the effect of the main pre-processing options on prediction of SSC and firmness; (2) identify informative wavelengths and relate them to the underlying chemistry; (3) compare the performance of the PLSR with wavelength-selected, kernel, tree-based and neural models; (4) apply the main PLSR to pixel-wise quality maps from hyperspectral images; (5) classify maturity stage and estimate storage time; and (6) use the main PLSR to convert the spectral predictions to a shelf-life estimate by means of first-order kinetics.

II. MATERIALS AND METHODS

Overall Framework

The framework shown in **Fig. 1** has six stages. Fruit is sampled throughout storage and measured spectrally; reference attributes are destructively measured on the same fruit; spectra are pre-processed; regression and classification models are calibrated and validated; and the results are used to interpret grading, ripening and shelf-life decisions. The calibration is updated by a model-maintenance loop, the most important condition for the long-term use of the spectral models [7].

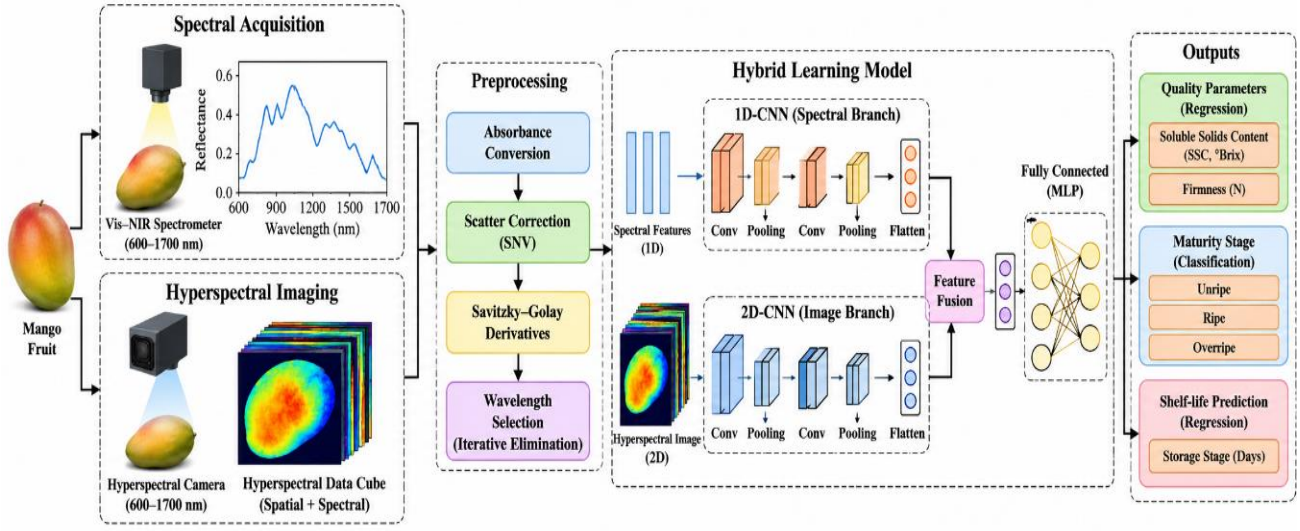


Fig 1. Framework for Intelligent Post-Harvest Quality Prediction from Vis–NIR Spectra and Hyperspectral Images.

Fruit Samples and Storage

The number of mango fruits used in the experimental design is 240 of commercial maturity from [orchard/location] to get the experimental results. The fruit was brought within 4 h to the laboratory where it was sorted for uniformity in size and freedom from defects and was stored at ambient temperature (25 ± 1 °C) at 70–75 %RH which is representative of marketing at ambient temperatures in the tropics. Fruits were randomly collected on eight sampling days (0, 2, 4, …, 14) which included both hard and green, and overripe stages. A range of days (not just one) of fruit storage means that the calibration is over a range of physiological variation encountered during the supply chain [1].

Spectral Acquisition
Diffuse reflectance spectra were acquired from two opposite equatorial positions of each fruit with a Vis–NIR spectrometer covering 600–1700 nm at 5 nm intervals (221 variables; [instrument model]), using a fibre-optic reflectance probe held at a fixed distance and angle against the skin in a dark enclosure. Each spectrum was the average of 32 scans and was referenced to a white reflectance standard and a dark current scan,

$$R = \frac{I_s - I_d}{I_w - I_d} \tag{1}$$

where I_s, I_w and I_d are the intensities from the sample, white reference and dark scan. The two spectra were averaged per fruit. For the imaging experiment, a line-scan hyperspectral camera covering the same range was used to image three representative fruit at storage days 0, 8 and 14, and a 64 × 64 pixel region containing the fruit was analysed [6].

Reference Measurements
Immediately after scanning, the reference attributes were measured on the scanned area. Firmness was measured as the maximum force (N) of a puncture test with an 8-mm cylindrical probe penetrating 10 mm at 2 mm s^{-1} in a texture analyser. SSC (°Brix) was determined with a temperature-compensated digital refractometer on juice from the same region. Storage time was recorded in days. Reference-method error, which sets a floor on the achievable prediction error, was taken as about 0.65 °Brix for SSC and 3.5 N for firmness in the demonstration data.

Spectral Pre-Processing
Reflectance was converted to apparent absorbance, which is approximately linear in concentration over the working range (Beer–Lambert law):

$$A = log_{10}\left(\frac{1}{R}\right) \tag{2}$$

Multiplicative and additive scatter effects, caused by differences in skin texture, flesh structure and probe contact, were removed by SNV [11], in which each spectrum is centred and scaled by its own mean $\bar{x}$ and standard deviation over the m wavelengths:

$$x_{SNV,j} = \frac{x_j - \bar{x}}{\sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(x_j - \bar{x})^2}} \tag{3}$$

Noise was reduced and overlapping bands resolved by Savitzky–Golay convolution smoothing and differentiation [10], which fits a local polynomial in a moving window of $2k + 1$ points and evaluates it (or its derivative) at the centre point:

$$x^*(j) = \frac{1}{N}\sum_{h=-k}^{k} c_h x_{j+h} \tag{4}$$

where c_h are the convolution coefficients and N is a normalising factor. A window of 15 points with a second-order polynomial was used. Five treatments were compared: raw absorbance, SG smoothing, SNV, SG first derivative, and SNV followed by SG first derivative.

Data Splitting and PLSR Calibration
The 240 fruit were divided into a calibration set (160) and an independent prediction set (80) by the Kennard–Stone algorithm applied to the leading principal-component scores of the spectra [12], which ensures that the calibration set covers the full spectral space. PLSR [13] projects the spectra onto a small number of latent variables that maximise covariance with the response,

$$X = TP^T + E, \quad y = Tq + f \tag{5}$$

where T are the scores, P and q the loadings, and E and f the residuals. Predictions follow from the regression vector b,

$$\hat{y} = Xb, \quad b = W(P^T W)^{-1} q \tag{6}$$

with W the weight matrix. The number of latent variables was chosen as the value that minimised the root-mean-square error of 10-fold cross-validation (RMSECV) within the calibration set.

Wavelength Selection
Informative wavelengths were selected with a simplified competitive adaptive reweighted sampling procedure [14]. In each of N = 60 iterations a random 80 % of the calibration samples was used to fit a PLSR model, and the variables with the

smallest absolute regression coefficients were removed so that the number retained followed an exponentially decreasing function,

$$p_i = pe^{-ki}, \quad k = \frac{ln\left(\frac{p}{p_{min}}\right)}{N-1} \tag{7}$$

where $p = 221$ is the number of original variables and $p_{min} = 12$. The subset with the lowest five-fold RMSECV was retained. This deterministic version omits the adaptive reweighted sampling of the original method but captures its principle of progressive coefficient-based elimination.

Machine-Learning Regressors

Three non-linear models were compared with PLSR using the SNV + SG first-derivative spectra. Support vector regression [15] used a radial basis function kernel,

$$K(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2) \tag{8}$$

with the penalty $C \in \{1, 10, 100\}$ and $\gamma \in \{10^{-4}, 3 \times 10^{-4}, 10^{-3}\}$ chosen by five-fold cross-validation. A random forest [16] of 300 trees and a multilayer perceptron with two hidden layers (64 and 32 neurons, L2 penalty 0.1, early stopping) completed the comparison. The neural network is a shallow stand-in for the deep convolutional models discussed in [8], which need much larger data sets.

Maturity Classification and Shelf-Life Kinetics

Fruit were assigned to three maturity stages according to storage time: unripe (days 0–4), ripe (days 6–10) and over-ripe (days 12–14). A support vector classifier was trained on the first six latent-variable scores of a PLSR model of storage time. Storage time itself was also predicted by PLSR. To convert predictions into a shelf-life estimate, the decline in firmness F with storage time t was described by first-order kinetics with a non-zero asymptote [17],

$$F(t) = F_\infty + (F_0 - F_\infty)e^{-kt} \tag{9}$$

where F_0 is the initial firmness, F_∞ the asymptotic firmness and k the rate constant (d^{-1}). The time t_{th} at which firmness reaches a threshold F_{th} (assumed 15 N for ready-to-eat fruit) is

$$t_{th} = \frac{-1}{k} ln\left(\frac{F_{th} - F_\infty}{F_0 - F_\infty}\right) \tag{10}$$

The kinetic model was fitted separately to the measured and the spectrally predicted mean firmness at each sampling day.

Performance Metrics and Pixel-Wise Mapping

Models were evaluated by the coefficient of determination, the root-mean-square error and the ratio of performance to deviation (RPD). With y_i the measured and $\hat{y}_i$ the predicted value for n samples,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad RPD = \frac{SD}{RMSEP} \tag{12}$$

These are reported as R^2_C and RMSEC for the calibration set and R^2_P, RMSEP and RPD for the prediction set; SD is the standard deviation of the reference values in the prediction set. By a widely used rule of thumb in near-infrared work, an RPD between about 2.5 and 3 permits screening, and values above 3 are adequate for quality control. For hyperspectral images, the calibrated PLSR model was applied to every pixel spectrum after the same pre-processing, and the resulting maps were smoothed by a 5 × 5 median filter.

III. RESULTS AND DISCUSSION

Spectral Characteristics and Reference Data

Fig. 2a shows the raw absorbance spectra coloured by storage time. Five main features are visible and their typical assignments are grouped together in **Table 1** [9]. The absorption at 680 nm (chlorophyll) was highest in fresh fruit but decreased with storage time, as the skin and flesh de-greened. The bands at ~970, ~1200 and ~1450 nm are due to O–H vibrations of water, of which the first overtone at 1450 nm is by far the strongest; the band at ~1200 nm is also due to a C–H second overtone of sugars and other carbohydrates, and the shoulder near 1690 nm is a C–H first overtone. The raw

spectra differ primarily in baseline and in the weak features that contain the information of interest; sugars are absorbed with much lower intensity than the dominant water bands and fruit absorbs and scatters light in different ways depending on firmness. Following SNV and SG differentiation (**Fig. 2b**), baseline differences were removed and bands were resolved; there was a systematic ordering of the curves by storage time around 680 nm and 1450 nm.

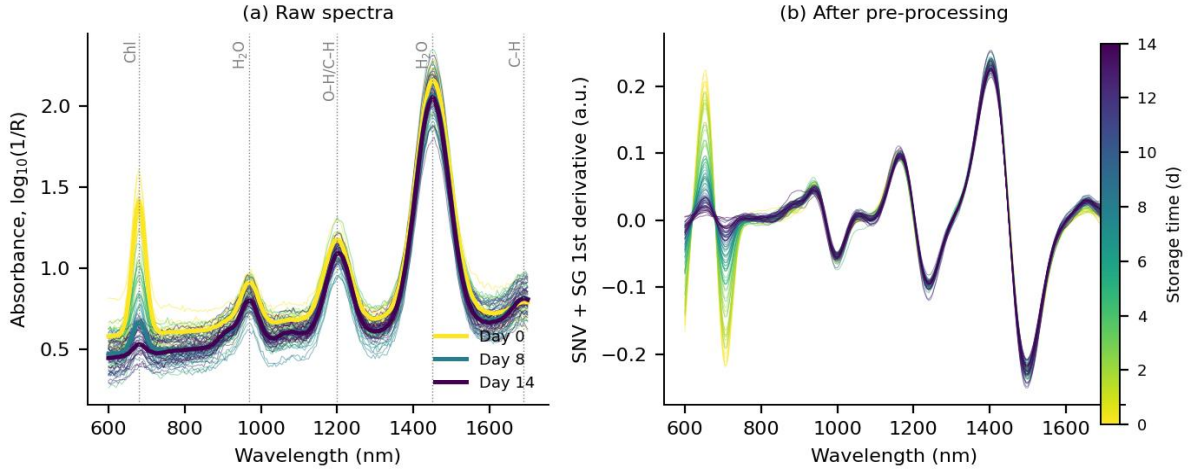


Fig 2. Simulated Vis–NIR Absorbance Spectra of Mango at Different Storage Times: (A) Raw Absorbance with The Main Absorption Bands Marked and The Mean Spectra of Days 0, 8 and 14 Highlighted; (B) The Same Spectra After SNV and Savitzky–Golay First Derivative.

Table 1. Main Absorption Bands in the 600–1700 Nm Range and Their Usual Assignments [9]

Wavelength (nm)	Assignment	Relevance to fruit quality
≈ 680	Chlorophyll a electronic absorption	Maturity and de-greening; decreases during ripening
≈ 910	C–H third overtone	Sugars and other carbohydrates
≈ 970	O–H second overtone	Water content and tissue state
≈ 1200	C–H / O–H second overtone	Sugars, water; texture-related scatter
≈ 1450	O–H first overtone	Water; strongest band, sensitive to cell-wall breakdown
≈ 1690	C–H first overtone	Sugars and organic acids

Table 2 summarises the reference data. The SSC increased from approximately 9 °Brix in fresh fruit to approximately 18 °Brix in over-ripe fruit, and the firmness decreased from approximately 60 N to less than 10 N, so as to span the whole commercial range. The standard deviations for the calibration set were 3.4 °Brix and 18.9 N respectively, and the mean SSC and firmness values for the prediction set were similar but slightly higher and lower respectively, which is a realistic feature of a Kennard–Stone splitting as extreme samples will be the ones that go into the calibration set.

Table 2. Descriptive Statistics of the Reference Attributes in the Calibration and Prediction Sets

Attribute	Set	Min	Max	Mean	SD
SSC (°Brix)	Calibration	6.3	19.4	14.3	3.4
	Prediction	7.0	19.0	15.2	3.1
Firmness (N)	Calibration	1.5	67.1	25.8	18.9
	Prediction	1.5	64.7	20.6	16.6
Storage time (d)	Calibration	0.0	14.0	6.7	4.8
	Prediction	0.0	14.0	7.7	4.2

Principal Component Analysis

Principal component analysis of the pre-processed spectra gave a first component that explained 91.1 % of the variance and a second that explained 3.2 % (**Fig. 3a**). The scores were a continuous continuum from fresh to overripe fruit and PC1 had a strong correlation with storage time (r = 0.89), indicating that ripening is the main source of variation in the spectra.

The loads were mainly aggregated around the chlorophyll band at 680 nm and the water band at 1450 nm, reflecting the softening process which includes the loss of green colour and the variation of water binding in the tissues (**Fig. 3b**). Scores for adjacent days were broadly similar and the gap between the first and last scores was quite small, even after day 8 which is where the adjacent maturity class confusions were reported in Section 3.8. This is to be expected because fruit which is a few days different in maturity at harvest can be at the same physiological age at the same time in storage.

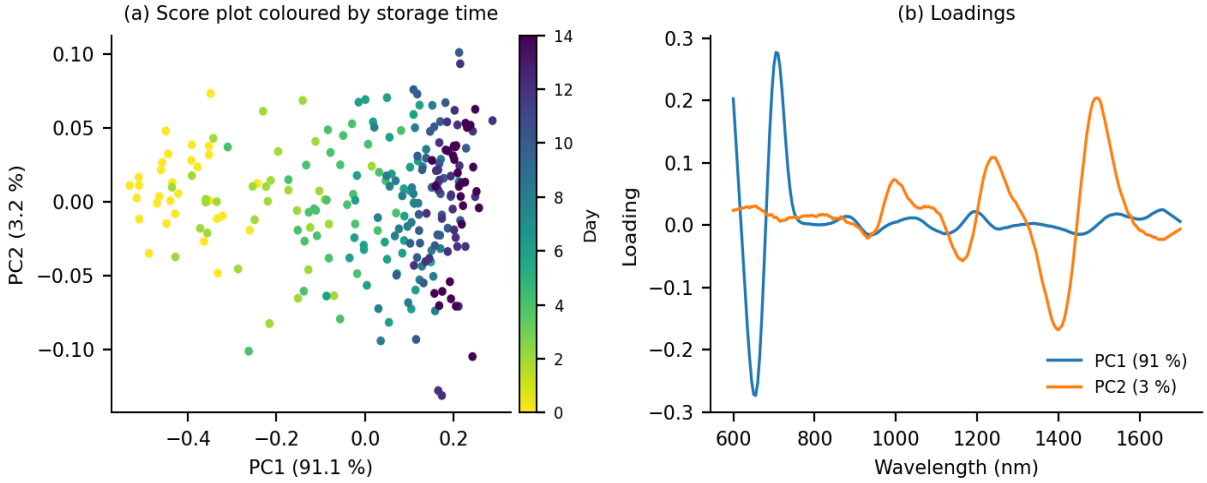


Fig 3. Principal Component Analysis of The Pre-Processed Spectra: (A) Score Plot Coloured by Storage Time and (B) Loadings of The First Two Components.

Effect of Pre-Processing

Table 3 shows the comparison of five pre-processing treatments with PLSR. The best prediction for SSC was achieved by SNV followed by the SG first derivative (RMSEP = 0.93 °Brix, R2P = 0.91), while the SG first derivative alone was the second (0.94 °Brix). The best result was obtained with the raw and smoothed absorbance (RMSEP: 0.98 °Brix) and the poorest with only SNV (RMSEP: 1.03 °Brix). The differences are small, 5-10 %, and the same trend was observed for firmness, with the lowest RMSEP value of 4.6 N obtained by SNV + SG first derivative and the highest value of 4.8 N obtained by SNV alone. There are two points that need to be emphasized. Derivatives caused no loss of amplitude information and reduced the number of latent variables required for firmness, to begin with. Secondly, the models were slightly degraded by SNV; by normalising each spectrum by its own standard deviation, part of the true amplitude information contained in each spectrum, which correlates with firmness due to scattering, is also lost. Pre-processing should be determined by cross validation on each new data set and not as a rule, as a few percent can still fall within the noise of a single data split. All further analyzes were based on the SNV + SG first-derivative representation.

Table 3. PLSR Performance for SSC and Firmness with Different Spectral Pre-Processing Treatments (LV = Latent Variables; SSC Columns Refer To °Brix and Firmness Columns to N)

Treatment	LV	R^2_C	RMSEC	R^2_P	RMSEP	RPD	RMSEP (F)	RPD (F)
Raw absorbance	4	0.94	0.86	0.90	0.98	3.2	4.7	3.6
SG smoothing	4	0.94	0.86	0.90	0.98	3.2	4.7	3.6
SNV	3	0.93	0.91	0.89	1.03	3.0	4.8	3.4
SG 1st derivative	6	0.95	0.73	0.91	0.94	3.3	4.7	3.5
SNV + SG 1st derivative	5	0.95	0.79	0.91	0.93	3.3	4.6	3.6

Wavelength Selection

The 221 variables were eliminated one by one, and the RMSECV was decreased step by step until approximately iteration 37, then it was leveled off and slightly increased as some informative variables were removed (**Fig. 4a**). The optimum subset consisted of 36 wavelengths (approximately 16 % of the spectrum) (**Fig. 4b**). They were not scattered randomly but clustered in the chemically meaningful regions of Table 1: 4 wavelengths in the chlorophyll region, 7 near 970 nm, 5 near 1200 nm, 12 around the 1450 nm water band and 3 beyond 1600 nm. The concentration of selected variables in the water bands, even for SSC, indicates that the sugar signal is in part indirect because of changes that occur in the water state and tissue structure as the fruit ripens; this general food safety warning applies to models of this type as they are likely to be based on correlated rather than causal features. This is important for robustness because if the calibration is based on an

indirect correlation, then it will not be robust if the relationship between sugar and water changes (such as between seasons and after store at a different temperature).

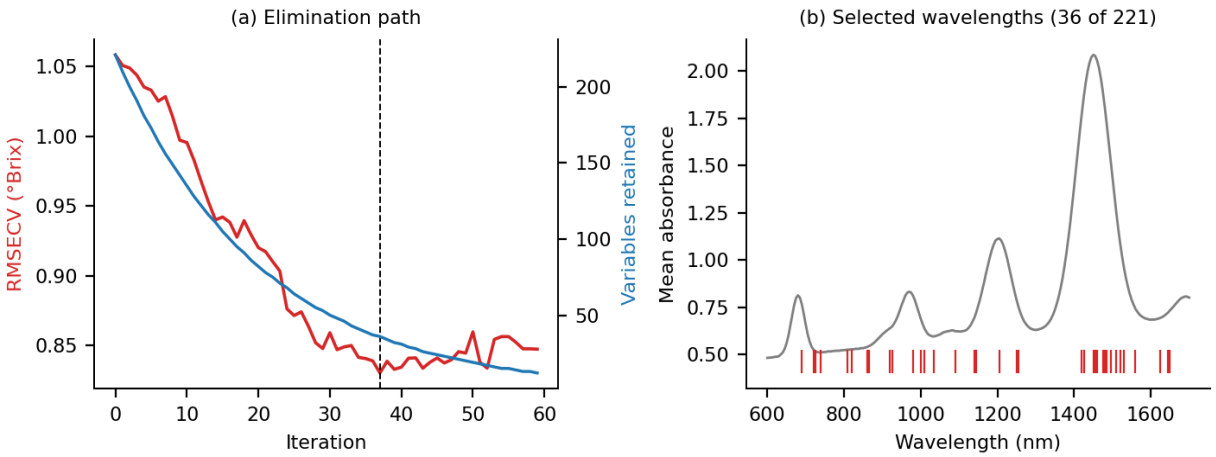


Fig 4. Wavelength Selection for SSC: (a) RMSECV and Number of Retained Variables During the Elimination (Dashed Line: Optimum) and (B) Selected Wavelengths (Red Ticks) On the Mean Absorbance Spectrum.

Calibration and Independent Prediction

Fig. 5 presents the measured against predicted values of the best PLSR models. For SSC the full-spectrum model (5 latent variables) yielded an R2C of 0.95 and RMSEC of 0.79 °Brix in the calibration set and R2P of 0.91, RMSEP of 0.93 °Brix and RPD of 3.3 in the independent prediction set. The model including 3 latent variables had the highest R2P (0.92), RMSEP (4.6 N), and RPD (3.6). The RPD values were not much below the calibration, so there was no serious over-fitting and all values above 3 satisfy the "rule of thumb" for good-quality control. There are two additional points that can be made. The average error of the refractometer reading is assumed to be 0.65 °Brix, which is roughly 92% of the RMSEP of 0.93 °Brix; thus, the majority of the error is reference method error and not spectral. Similarly, the scatter of the data points around the identity line is wider in the hard end of the range and the spread of force values around the day of harvest maturity is also greater, as small differences in day of harvest maturity would result in large difference in force values at the hard end of the range. A prediction error of this size is satisfactory for grade purposes if only three to four grades are to be made but not at all satisfactory if differences of one or two °Brix between individual fruits are to be resolved.

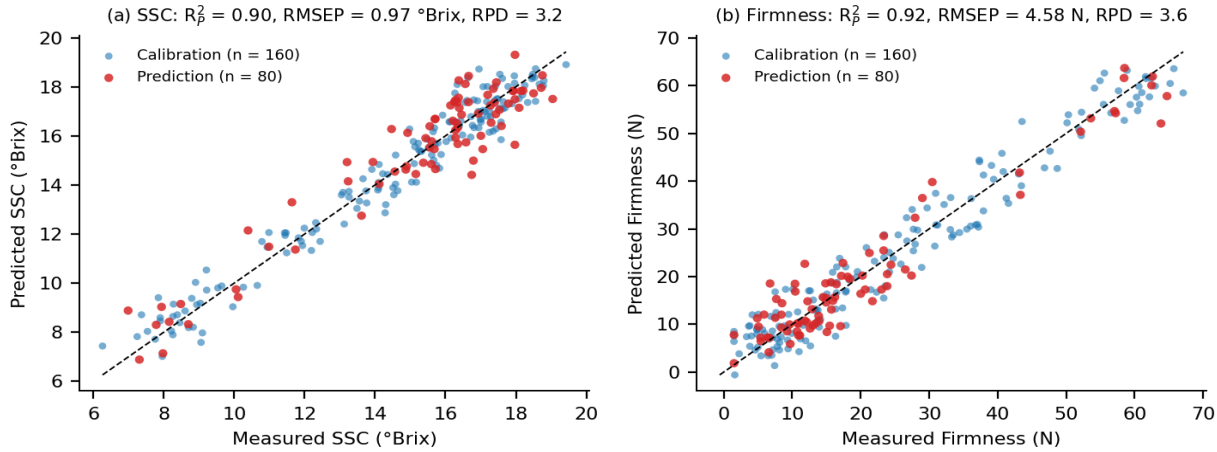


Fig 5. Measured Versus Predicted Values for (A) SSC and (B) Firmness Using PLSR on the Selected Wavelengths (Blue: Calibration Set; Red: Independent Prediction Set; Dashed Line: 1:1).

Comparison of Models

Table 4 and **Fig. 6** compare the five modelling approaches on the same prediction set. PLSR on the full spectrum gave the lowest SSC error (RMSEP = 0.93 °Brix), with the support vector regressor (0.96 °Brix) and wavelength-selected PLSR (0.97 °Brix) very close behind and the random forest slightly worse (1.01 °Brix). For firmness the support vector regressor was marginally best (4.47 N) ahead of PLSR (4.56 N), but the difference is far smaller than the uncertainty of the reference measurement and should not be interpreted as a real advantage. Reducing the spectrum to 36 wavelengths cost about 4 % in SSC accuracy and almost nothing in firmness accuracy (4.58 N), which is encouraging for the design of inexpensive

multispectral sensors using a few filters or LEDs in place of a full spectrometer. The multilayer perceptron was clearly the weakest (SSC RMSEP = 1.68 °Brix; firmness RMSEP = 6.3 N), because a network with several thousand weights is poorly constrained by 160 calibration spectra; this does not imply that deep models cannot work, only that they require much larger and more diverse data sets or strong regularisation, augmentation and transfer learning [8]. For small post-harvest data sets, linear latent-variable models therefore remain the sound baseline, and any more complex model must beat them under the same validation protocol.

Table 4. Performance of the Regression Models on The Independent Prediction Set (SSC in °Brix, Firmness in N)

Model	R^2_P (SSC)	RMSEP	RPD	R^2_P (F)	RMSEP	RPD
PLSR (full spectrum)	0.91	0.93	3.3	0.92	4.6	3.6
CARS-style + PLSR	0.90	0.97	3.2	0.92	4.6	3.6
SVR (RBF)	0.90	0.96	3.2	0.93	4.5	3.7
Random forest	0.89	1.01	3.1	0.90	5.2	3.2
MLP (64-32)	0.70	1.68	1.9	0.85	6.3	2.6

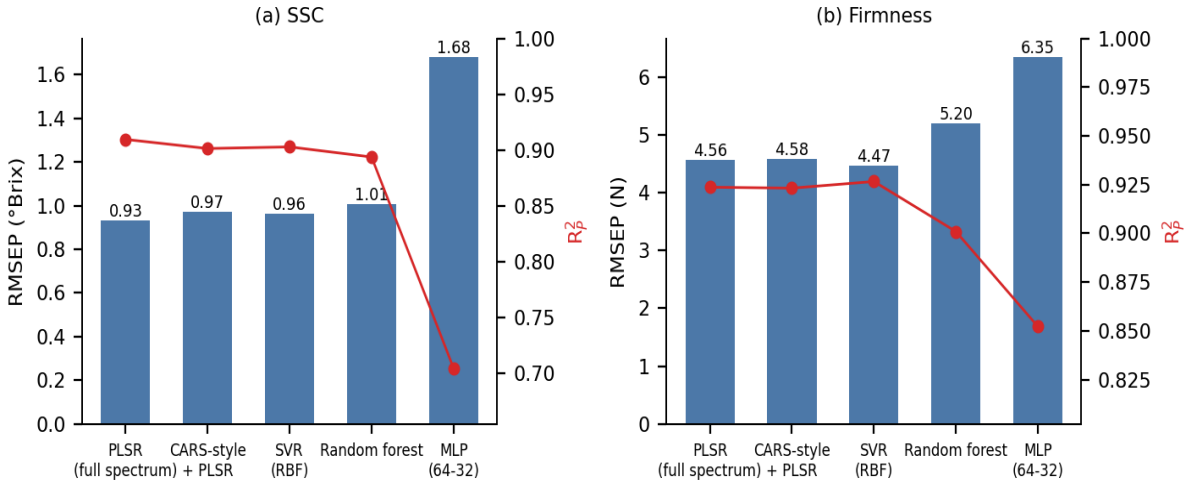


Fig 6. Comparison of Regression Models on The Prediction Set: RMSEP (Bars) and R^2_P (Red Markers) For (A) SSC and (B) Firmness.

Hyperspectral Mapping of Quality Within the Fruit

Point spectroscopy gives one value per measuring position, whereas hyperspectral imaging reveals how quality is distributed across the fruit. **Fig.7** shows pseudo-colour images and the pixel-wise SSC and firmness maps obtained by applying the calibrated models to 1609 pixel spectra of a fruit at days 0, 8 and 14. The maps reproduce the expected evolution: the predicted mean SSC rose from 9.2 °Brix on day 0 through 14.1 °Brix on day 8 to 17.9 °Brix on day 14, and the mean predicted firmness fell from 53.7 N to 27.0 N and 5.6 N. At the intermediate stage the maps show a gradient from the stalk end to the apex, which mirrors the way mango ripens unevenly from the stalk end and cannot be captured by a single-point reading. Against the simulated ground-truth maps, the pixel-wise agreement was R^2 = 0.99 (RMSE = 0.26 °Brix) for SSC and R^2 = 0.95 (RMSE = 4.2 N) for firmness. The pixel-level SSC error is lower than the RMSEP of the fruit-level model because the maps are compared with the true simulated values rather than with noisy refractometer readings and because median filtering suppresses random pixel noise; with real data the error will be larger. Firmness at day 8 was overestimated by about 7.0 N, which indicates that the model is least reliable in the steepest part of the softening curve. Such maps support the detection of internal bruising, uneven ripening and chilling injury, and give breeders and packhouse managers a quantitative basis for choosing ripening treatments, although the camera and processing cost is still higher than that of point sensors.

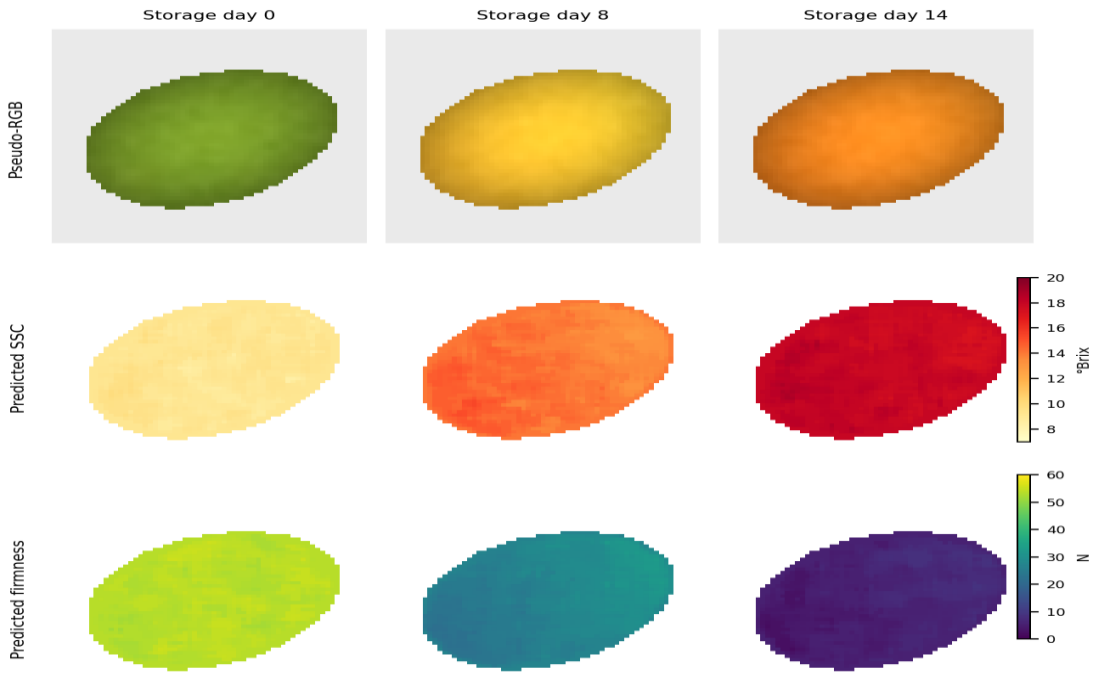


Fig 7. Simulated Hyperspectral Results for a Fruit at Storage Days 0, 8 and 14: Pseudo-RGB Image (Top), Pixel-Wise Predicted SSC (Middle) and Predicted Firmness (Bottom).

Maturity Classification and Storage-Time Prediction

Fig. 8a shows the confusion matrix of the support vector classifier for the three maturity stages on the prediction set. The overall accuracy was 0.75 (60 of 80 fruit), with class recalls of 0.76 for unripe, 0.72 for ripe and 0.80 for over-ripe fruit. All errors were between neighbouring classes except one over-ripe fruit classified as unripe, which is the pattern expected for a continuous ripening process in which class boundaries are arbitrary. The middle class was most often confused with the over-ripe class, which is of practical importance because the transition from ripe to over-ripe determines whether fruit can still be transported. Storage time was predicted with $R^2_P = 0.83$, RMSEP = 1.71 d and RPD = 2.5 (**Fig. 8b**). The error of about 2 to 2 days is a consequence of batch variability in harvest maturity rather than of poor spectral measurement, since fruit of the same age differ in their physiological stage. For this reason, storage time is better treated as a proxy for stage than as a calendar quantity, and the quality attributes SSC and firmness, which are physically linked to eating quality, are the more useful targets for decision-making.

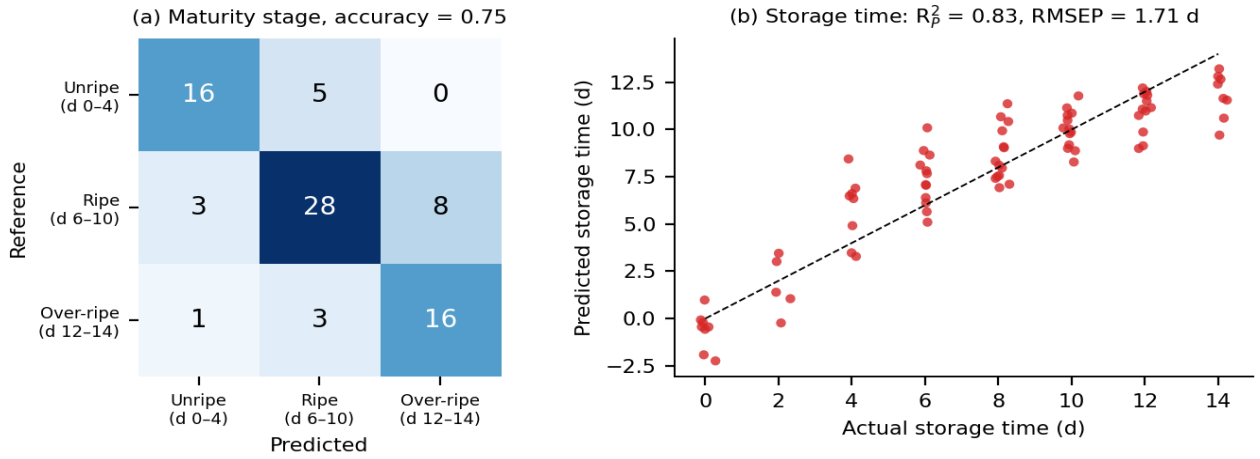


Fig 8. (a) Confusion Matrix for Classification of Maturity Stage and (B) Predicted Versus Actual Storage Time on The Prediction Set.

Shelf-Life Prediction

Fig. 9 shows the mean firmness at each sampling day, both measured and predicted from the spectra, together with fitted first-order kinetic curves (Eq. 11). The measured data gave $F_0 = 59.2$ N, $F_\infty = 2.8$ N and $k = 0.189$ d^{-1}, whereas the spectrally

predicted firmness gave F_0 = 58.3 N, F_∞ = 2.2 N and k = 0.179 d^{-1}. The two curves are almost indistinguishable. With a ready-to-eat threshold of 15 N, Eq. (12) gives a threshold time of 8.1 d from the measured data and 8.2 d from the spectral predictions, a difference of about 3 h. Agreement at the level of batch averages is much better than at the level of individual fruit, because random errors cancel in the mean, so the approach is suited to lot-level decisions such as when to dispatch a consignment or open a ripening room, but not to guaranteeing the shelf life of one individual fruit. Because the rate constant depends strongly on temperature, the kinetic parameters estimated at 25 °C must be recalibrated for cold-chain conditions, for example by an Arrhenius-type temperature dependence.

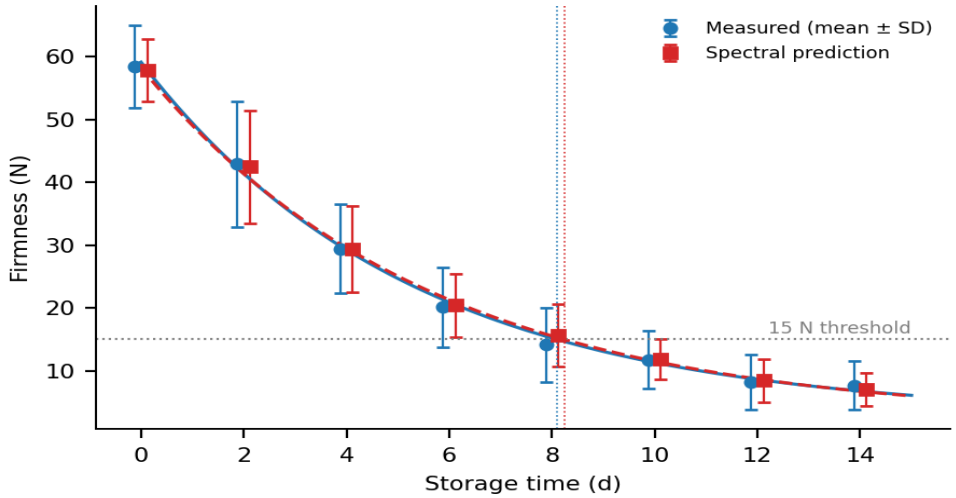


Fig 9. Firmness Decline During Storage: Measured (Blue) and Spectrally Predicted (Red) Means ± SD With First-Order Kinetic Fits; Vertical Lines Mark the Time at Which The 15 N Threshold Is Reached.

Implications For Post-Harvest Management

The results suggest several concrete uses (**Table 5**). On a packhouse line, a point Vis–NIR sensor predicting SSC and firmness at a few dozen milliseconds per fruit allows grading by eating quality rather than by size and colour alone, so that premium fruit can be directed to distant or high-value markets and soft fruit to local or processing channels. In ripening rooms, repeated spectral scans of sample fruit can replace destructive sampling to decide when ethylene treatment or cooling should be started or ended, saving energy and reducing over-ripening. In the supply chain, lot-level shelf-life estimates support dynamic stock rotation, in which batches with the shortest remaining life are sold first, reducing waste. Finally, hyperspectral maps can serve as a research and quality-assurance tool. These uses contribute to the broader aims of cutting post-harvest loss and improving farm income, and the non-destructive nature of the method means that every fruit, rather than a small sample, can in principle be assessed [1,3].

Table 5. Suggested Deployment Scenarios for Spectral Quality Prediction

Scenario	Sensor and output	Decision supported
Packhouse grading	Point Vis–NIR; SSC, firmness grade	Sort by eating quality; assign market and price class
Ripening room	Sample scans; stage class	Start or stop ethylene and cooling; avoid over-ripening
Supply chain	Lot averages; shelf-life estimate	Dispatch order, stock rotation, markdown timing
Research and QA	Hyperspectral imaging; quality maps	Detect uneven ripening, bruising and injury; test treatments

Limitations and Future Work

Several limitations apply. First, the numerical results are simulated and must be verified with measured data from several cultivars, orchards, seasons and instruments. Second, models calibrated with one instrument or season often lose accuracy when applied to another, and calibration transfer, model updating and the inclusion of the variability in the calibration set are essential before deployment [7]. Third, light penetrates only a few millimetres into the fruit flesh, so a measurement of the skin side may not represent the core, and the choice of geometry and of the measurement position matters [5]. Fourth, temperature of the fruit, skin colour, and surface moisture affect the spectrum and should be controlled or included in the calibration. Fifth, the prediction error of about 0.9 °Brix and 4.6 N, which is partly reference-method error, is adequate for grading but not for resolving small differences between fruit. Future work will collect real multi-season data, test

calibration-transfer methods, evaluate low-cost multispectral sensors using the selected wavelengths, and examine deep learning with data augmentation and transfer learning [8] once larger data sets are available.

IV. CONCLUSION

A framework combining Vis–NIR spectroscopy, spectral pre-processing, PLSR and machine-learning models can predict the internal quality and storage stage of fruit without damaging it. In the demonstration data set, PLSR on SNV + SG first-derivative spectra predicted SSC and firmness with RPD values of 3.3 and 3.6, and a selection of 36 of 221 wavelengths in the water and sugar regions retained most of the accuracy. A support vector regressor matched PLSR, whereas a small neural network did not, so that linear latent-variable models remain the robust baseline for small data sets. Application of the calibrations to hyperspectral images produced quality maps that reveal within-fruit gradients, the maturity class was predicted with 75 % accuracy, and a first-order kinetic model fed by spectral predictions estimated the time to reach 15 N firmness to within about 3 h of the measured value. These capabilities allow quality-based grading, ripening-room control and shelf-life-based logistics. The next steps are validation on measured multi-cultivar and multi-season data, calibration transfer between instruments, and the development of low-cost multispectral devices for use on farms and in packhouses.

CRediT Author Statement

The author reviewed the results and approved the final version of the manuscript.
Conceptualization: Lise Roy Porter and Mekdes Solomon Cherinet; **Methodology:** Lise Roy Porter and Mekdes Solomon Cherinet; **Software:** Lise Roy Porter and Mekdes Solomon Cherinet; **Data Curation:** Lise Roy Porter and Mekdes Solomon Cherinet; **Writing- Original Draft Preparation:** Lise Roy Porter and Mekdes Solomon Cherinet; **Visualization:** Lise Roy Porter; **Investigation:** Lise Roy Porter and Mekdes Solomon Cherinet; **Supervision:** Lise Roy Porter; **Validation:** Mekdes Solomon Cherinet; **Writing- Reviewing and Editing:** Mekdes Solomon Cherinet; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received for conducting this research.

Competing Interests

There are no competing interests.

References

1. M. Palumbo et al., “Emerging Postharvest Technologies to Enhance the Shelf-Life of Fruit and Vegetables: An Overview,” Foods, vol. 11, no. 23, p. 3925, Dec. 2022, doi: 10.3390/foods11233925.
2. S. Tian and H. Xu, “Nondestructive Methods for the Quality Assessment of Fruits and Vegetables Considering Their Physical and Biological Variability,” Food Engineering Reviews, vol. 14, no. 3, pp. 380–407, Jan. 2022, doi: 10.1007/s12393-021-09300-0.
3. K. Chi, J. Lin, M. Chen, J. Chen, Y. Chen, and T. Pan, “Changeable moving window-standard normal variable transformation for visible-NIR spectroscopic analyses,” Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 308, p. 123726, Mar. 2024, doi: 10.1016/j.saa.2023.123726.
4. S. L. McGuirk and I. H. Cairns, “Relationships between Soil Moisture and Visible–NIR Soil Reflectance: A Review Presenting New Analyses and Data to Fill the Gaps,” Geotechnics, vol. 4, no. 1, pp. 78–108, Jan. 2024, doi: 10.3390/geotechnics4010005.
5. Y. Bu, S. Li, C. Gong, and W. Guo, “Optical transmission simulation-guided optimization of Vis/NIR detection and simultaneous multi-quality prediction for post-ripening kiwifruit,” Food Control, vol. 191, p. 112487, Jan. 2027, doi: 10.1016/j.foodcont.2026.112487.
6. M. Nikzadfar et al., “Hyperspectral Imaging Aiding Artificial Intelligence: A Reliable Approach for Food Qualification and Safety,” Applied Sciences, vol. 14, no. 21, p. 9821, Oct. 2024, doi: 10.3390/app14219821.
7. W. Saeys, N. Nguyen Do Trong, R. Van Beers, and B. M. Nicolaï, “Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review,” Postharvest Biology and Technology, vol. 158, p. 110981, Dec. 2019, doi: 10.1016/j.postharvbio.2019.110981.
8. X. Zhang, T. Lin, J. Xu, X. Luo, and Y. Ying, “DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis,” Analytica Chimica Acta, vol. 1058, pp. 48–57, Jun. 2019, doi: 10.1016/j.aca.2019.01.002.
9. H. Yan, M. D. G. Neves, B. M. Wise, I. A. Moraes, D. F. Barbin, and H. W. Siesler, “The Application of Handheld Near-Infrared Spectroscopy and Raman Spectroscopic Imaging for the Identification and Quality Control of Food Products,” Molecules, vol. 28, no. 23, p. 7891, Dec. 2023, doi: 10.3390/molecules28237891.
10. C. Yan, “A review on spectral data preprocessing techniques for machine learning and quantitative analysis,” iScience, vol. 28, no. 7, p. 112759, Jul. 2025, doi: 10.1016/j.isci.2025.112759.
11. J. Cao and H. Yang, “A dynamic normalized difference index for estimating soil organic matter concentration using visible and near-infrared spectroscopy,” Ecological Indicators, vol. 147, p. 110037, Mar. 2023, doi: 10.1016/j.ecolind.2023.110037.
12. K. Fu, S. G. Vougioukas, and B. N. Bailey, “Computer-aided design and optimization of a multi-level fruit catching system for fresh-market fruit harvesting,” Computers and Electronics in Agriculture, vol. 225, p. 109334, Oct. 2024, doi: 10.1016/j.compag.2024.109334.
13. C. Dias, T. R. S. Brandão, C. Cassoni, M. W. Vasconcelos, A. Ferrante, and M. Pintado, “Firmness prediction of ‘Rocha’ pear based on non-destructive methods and shelf-life using PLS-regression modeling,” Postharvest Biology and Technology, vol. 205, p. 112493, Nov. 2023, doi: 10.1016/j.postharvbio.2023.112493.
14. Y. Li and X. Yang, “Quantitative analysis of near infrared spectroscopic data based on dual-band transformation and competitive adaptive reweighted sampling,” Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 285, p. 121924, Jan. 2023, doi: 10.1016/j.saa.2022.121924.

15. D. Satyaraj, T. Prabakaran, Y. M. Blessy, L. N. Jayanthi, S. K. Mouleeswaran, and S. Velmurugan, “Enhancing Fruit and Vegetable Preservation with Support Vector Machine and IoT Connectivity,” 2024 International Conference on Automation and Computation (AUTOCOM), pp. 237–242, Mar. 2024, doi: 10.1109/autocom60220.2024.10486142.
16. B. R. Ramos Collin et al., “Random forest regressor applied in prediction of percentages of calibers in mango production,” Information Processing in Agriculture, vol. 12, no. 3, pp. 370–383, Sep. 2025, doi: 10.1016/j.inpa.2024.12.002.
17. N. Brahimi et al., “Evaluation and validation criteria for kinetic models in food science: A critical review,” Trends in Food Science & Technology, vol. 156, p. 104857, Feb. 2025, doi: 10.1016/j.tifs.2024.104857.

Publisher’s note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.

ISSN: 3104-4654