Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

https://doi.org/10.17113/ftb.64.04.26.9553 original scientific paper

Cross-Market Non-Destructive Quality Assessment of Arumanis Mangoes Using Genetic Algorithm–Optimized Ensemble Models and Low-Cost AS7265x UV–Vis–NIR Spectroscopy

Running title: Non-Destructive Quality Assessment of Mango

Susanto B. Sulistyo[1]*, Desta L. Fatmala[1], Dini Y. Putri[1], Muhammad R. Nursyahpratama[1], Krissandi Wijaya[1], Purwoko H. Kuncoro[1], Arief Sudarmaji[1], Pepita Haryanti[2] and Hasan M. Kadhim[3]

[1]Department of Agricultural Engineering, Faculty of Agriculture, Jenderal Soedirman University, Jl. Dr. Soeparno 63, Purwokerto 53122, Indonesia
[2]Department of Food Technology, Faculty of Agriculture, Jenderal Soedirman University, Jl. Dr. Soeparno 63, Purwokerto 53122, Indonesia
[3]Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Abi Talib Street, Baghdad 10052, Iraq

Received: 23 December 2025
Accepted: 8 November 2026

Copyright © 2026 Authors retain copyright and grant the FTB journal the right of first publication under CC-BY 4.0 licence that allows others to share the work with an acknowledgement of the work's authorship and initial publication in the journal

SUMMARY

Research background. Arumanis mango is a high-value tropical fruit whose market acceptance depends primarily on internal quality attributes. However, reliable quality grading remains difficult because Arumanis is a green-mature cultivar that shows minimal external colour changes during ripening. Moreover, most existing non-destructive sensing studies rely on controlled experimental conditions, high-cost instruments, or homogeneous sample sources, with limited evaluation of model robustness across heterogeneous market environments.

Experimental approach. This research evaluates a low-cost, portable, and non-destructive

*Corresponding author:
Phone: +6281225417015
E-mail: susanto.sulistyo@unsoed.ac.id

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

quality assessment framework combining an AS7265x UV-Vis-NIR multispectral sensor with a genetic algorithm-optimised committee machine (GA-CM) to predict physicochemical parameters of Arumanis mangoes. A total of 200 samples were collected from both modern and traditional markets to capture variability in maturity levels and postharvest handling practices. Non-destructive spectral reflectance data were acquired at 18 wavelengths (410–940 nm), followed by destructive reference measurements of total soluble solids, total sugar, vitamin C, pH, and water content. Three base learners, *e.g.* linear regression (LR), support vector regression (SVR), and artificial neural network (ANN), were developed and integrated using two ensemble strategies: simple averaging and GA-based weighted averaging. Model performance was assessed using the coefficient of determination (R^2) and mean squared error (MSE).

Results and conclusions. The results demonstrate that mangoes from different market typologies exhibit similar spectral trends but differ in reflectance magnitude, reflecting variations in maturity and postharvest conditions and yielding a robust, heterogeneous dataset. Among individual models, ANN and SVR significantly outperformed LR, with ANN showing superior capability in capturing non-linear relationships between spectral signatures and internal quality attributes. Ensemble modelling further enhanced prediction accuracy, with the GA-CM consistently outperforming both individual learners and simple averaging by optimally weighting base models. High predictive performance was achieved across all quality parameters, with an average R^2 of approximately 0.97.

Novelty and scientific contribution. This study provides a novel contribution by introducing a GA-optimised ensemble framework that leverages complementary modelling strengths under low spectral resolution conditions. By validating the system across mixed supply chains rather than controlled environments, the proposed approach demonstrates strong generalisation capability and practical applicability for real-world, non-destructive quality assessment of Arumanis mango.

Keywords: mango quality assessment; multispectral sensor; genetic algorithm-optimised committee machine; artificial neural network; linear regression; support vector regression

INTRODUCTION

The mango (*Mangifera indica* L.) is one of the most economically significant tropical fruits globally, favoured for its exotic flavour and nutritional value. In the Southeast Asian market, the 'Arumanis' cultivar is particularly dominant due to its distinct aroma and high consumer demand. However, the commercial value of Arumanis mangoes is frequently compromised by inconsistencies

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

in post-harvest quality grading. While many other varieties show clear changes in colour as they ripen, the Arumanis mango is a 'green-mature' type, that retains its green skin even at full maturity, which results in a large mismatch between appearance and internal quality properties [*1*]. As a result, traditional methods of sorting the fruit by size or by the colour of the skin are generally not reliable in predicting important internal physicochemical properties such as water content, total soluble solids (TSS), total sugar content, acidity (pH), and vitamin C [*2*].

To address these quality control challenges, the agricultural industry has historically relied on destructive analytical methods, such as refractometry for TSS and titration for acidity. While accurate, these methods are time-consuming, labour-intensive, and result in the destruction of the sample, rendering them unsuitable for real-time grading in high-volume supply chains [*3*]. In response, non-destructive testing (NDT) technologies based on near-infrared (NIR) spectroscopy have emerged as a standard solution. Spectroscopy links light absorption at specific wavelengths to molecular bonds like C–H and O–H. This makes it possible to analyse a fruit's internal chemistry non-destructively, without needing to cut or damage it [*4*]. However, commercial NIR spectrometers are typically bulky, complex, and prohibitively expensive ($10,000–$50,000), restricting their adoption to large-scale industrial facilities and leaving smallholder farmers with limited access to precision grading technology.

The recent advent of Micro-Electro-Mechanical Systems (MEMS) has facilitated the development of low-cost, portable spectral sensors, such as the AS7265x triad sensor, which covers the ultraviolet, visible, and near-infrared (UV-Vis-NIR) spectrums (410–940 nm). These sensors are more affordable than full-scale lab equipment; however, they have low spectral resolution. Benchtop spectrometers capture hundreds of continuous light wavelengths, whereas these sensors usually rely on just 18 separate channels [*5*]. That limited view creates sparse, heavily overlapping data, making basic linear regression models fall short when predicting complex biochemical traits like vitamin C or specific sugars, which do not follow neat, straight-line patterns [*6*].

To overcome the inherent hardware limitations of low-cost sensors, recent computational advancements suggest shifting focus from hardware improvement to advanced algorithmic compensation. Ensemble learning, specifically the concept of committee machines, has gained traction in precision agriculture [*7*]. A Committee machine aggregates the predictions of multiple diverse expert systems or base learners, such as artificial neural networks (ANN), support vector regression (SVR), and linear regression (LR), to achieve generalization performance superior to any single model [*8*]. The underlying premise is that different algorithms capture different aspects of the

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

data structure; for instance, ANNs excel at capturing non-linear dependencies typical of biological variability [*9*], while LR is robust against overfitting in linear trends [*10*].

However, a standard committee machine typically employs simple averaging, which may not be optimal if one base learner significantly outperforms the others for a specific parameter. A more sophisticated approach involves dynamically optimizing the voting weights of the committee members. Genetic algorithms (GA), a stochastic search heuristic inspired by natural evolution, offer a robust mechanism to determine the optimal weight combination that minimizes the prediction error for specific target variables [*11*]. Despite the potential of this hybrid approach, there is a scarcity of literature applying GA-optimized committee machines specifically to the domain of low-cost spectral sensing for tropical fruit quality.

This study aims to bridge the gap between affordable hardware and high-precision analysis. We propose the development of a portable, non-destructive grading system using the AS7265x sensor, integrated with a genetic algorithm-optimized committee machine. The specific objective is to accurately estimate five key physicochemical quality parameters of Arumanis mangoes: total sugar, total soluble solids (TSS), vitamin C, pH, and water content. By leveraging the computational strength of ensemble learning to compensate for sensor resolution, this research seeks to validate a cost-effective solution for real-time fruit quality assessment in mixed-market supply chains.

MATERIALS AND METHODS

Sample preparation

A total of 200 'Arumanis' mango (*Mangifera indica* L.) samples were utilized in this study. To ensure the development of a robust calibration model capable of handling real-world supply chain heterogeneity, samples were stratified into two distinct acquisition groups. The first group (N_m=100) was sourced from modern retail outlets (supermarkets) in Purwokerto, Indonesia, characterized by controlled storage temperatures (20–22 °C), while the second group (N_t=100) was obtained from traditional open-air markets, located in Purwokerto, Indonesia, exposing the fruit to fluctuating ambient conditions (28–32 °C). This sampling strategy was designed to introduce maximum variability in physiological maturity and internal quality attributes, preventing model overfitting to specific storage conditions. All samples were equilibrated to laboratory temperature (25 °C) and cleaned prior to spectral acquisition to minimize surface interference.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Instrumentation and spectral acquisition

The spectral data acquisition was performed using a custom-designed prototype integrating the AS7265x smart spectral sensor triad (ams-OSRAM AG, Premstaetten, Austria) coupled with Arduino Uno microcontroller (Arduino, Monza, Italy) (Fig. S1). The AS7265x chipset includes three specialized sensors, *e.g.* AS72651, AS72652, and AS72653, to capture the spectral range from 410 to 940 nm. The device contains a built-in light source that includes a white LED at 5700 K, a UV LED at 405 nm, and an NIR LED at 875 nm. A plastic hardcase was utilized to cover the instrument so that the spectral responses could be measured. It was then connected to a computer to determine the sample's reflectance light via USB type A and B connections.

The measurement setup involved placing the fruit sample on a designated testing aperture, with the sensor module mounted directly underneath to facilitate non-destructive contact measurement (Fig. S2). Moreover, due to the diverse colour range of the Arumanis mango skin, spectral measurements were taken from two opposite areas on the surface of each fruit to account for variations in colour (Fig. S3). This method was employed to minimize the impact of small, localized colour changes on both surfaces, while still obtaining spectral information that can provide more precise depictions of how the fruit is overall.

Upon activation of the system, the integrated LED array emitted broad-spectrum radiation spanning ultraviolet (UV), visible (Vis), and near-infrared (NIR) bands onto the specific surface area of the sample. During this interaction, a portion of the light energy was absorbed by the fruit tissue, correlated with its internal physicochemical properties, while the remaining light was reflected. This diffuse reflectance was subsequently captured by the sensor’s three distinct optical units (UV, Vis, and NIR), which generated signals corresponding to the specific wavelengths detected.

These analogue spectral signals were processed and converted into digital format via an Arduino-based interface. The resulting data were systematically recorded and stored as a structured dataset for further analysis. The AS7265x sensor enabled the acquisition of high-fidelity reflectance intensities across 18 discrete wavelengths covering the UV-Vis-NIR range, providing a comprehensive optical profile essential for characterizing the fruit quality attributes.

Measurement of mango physical dimensions

Each mango sample was measured for its physical dimensions, *i.e.* mass, size, and surface area. Mango mass and size (length, width, and thickness) were measured twice using a weight balance and a caliper, respectively. Mango surface area was calculated by means of the Knud Thomsen ellipsoidal approximation using the following equation [*12*]:

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

$$A \approx 4\pi\left(\frac{(a\cdot b)^p+(a\cdot c)^p+(b\cdot c)^p}{3}\right)^{\frac{1}{p}} \qquad /1/$$

where *a*, *b*, and *c* were the radius of the fruit length axis, width axis, and thickness axis, respectively (Fig. S4), and *p*=1.6075.

Reference measurement of physicochemical properties

Immediately following non-destructive spectral scanning, destructive analyses were performed to determine the ground-truth values of five key mango quality parameters. All chemical analyses were conducted in triplicate. For the determination of total soluble solids (TSS), the mango pulp was homogenized and filtered, and the resulting supernatant was measured using a digital refractometer (model Atago PAL-1, Tokyo, Japan). The TSS content was expressed in degrees Brix (°Bx).

Total sugar content, which is distinct from TSS, was determined by quantifying the combined concentrations of sucrose, glucose, and fructose using the anthrone colorimetric method with spectrophotometric detection at 620 nm [*13*]. This parameter was included because sugar-related absorption bands in the NIR region (C-H stretch) are important spectral features for the model development.

The acidity of the mango samples was evaluated by measuring the pH of the juice using a calibrated digital ATC pocket pH meter (model Hofun pH-009, New Delhi, India) at 25 °C. Vitamin C (ascorbic acid) content was determined using the standard iodine titration method [*14*]. Approximately 10 g of mango pulp was homogenized with distilled water and subsequently filtered. The filtrate was titrated with a standardized 0.01 N Iodine (I_2) solution using starch as an indicator until a persistent blue colour was observed. Iodine (I_2) and soluble starch were purchased from Merck (Darmstadt, Germany). The vitamin C content (mg/100 g) was calculated using the following equation:

$$w(\text{vitamin C}) = \left(\frac{\left(V_\text{y}-V_\text{b}\right)\cdot c\cdot 88.06\cdot \text{DF}}{m_\text{p}}\right)\cdot 100 \qquad /2/$$

where V_y is the volume of iodine used for the sample (mL), V_b is the volume for the blank (mL), *c* is the molarity of the iodine solution (0.01 N), 88.06 is the equivalent mass of ascorbic acid, DF is the dilution factor, and m_p is the mango pulp mass (g).

Water content was determined gravimetrically in accordance with the AOAC official standards [*15*]. Approximately 5 g of fresh mango pulp sample was weighed and dried in a convection oven at

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

105 °C until a constant mass was achieved. The water content was expressed as a percentage (%) on a wet basis and calculated using the following equation:

$$w(\text{water}) = (\frac{m_1 - m_2}{m_1}) \cdot 100 \quad /3/$$

where m_1 and m_2 are the mass of the initial and dried sample (g), respectively.

Data processing and committee machine architecture

Base learners

All computational algorithms and model developments were implemented using the MATLAB R2020a programming environment [*16*]. The committee machine was constructed using three diverse base algorithms, each defined by specific mathematical formulations to model the relationship between the spectral input matrix, X (reflectance at 18 wavelengths), and the corresponding physicochemical quality parameters, Y. Three base learners employed in this study were linear regression, support vector regression, and artificial neural network.

Linear regression (LR) was employed as the baseline linear model to capture direct correlations. It models the target variable as a linear combination of the spectral features plus an error term. The predicted physicochemical quality parameter using LR can be obtained from the following equation [*17*]:

$$Y_{\text{LR}} = \beta_0 + \sum_{\text{i}=1}^{18} (\beta_{\text{i}} \cdot X_{\text{i}}) + \epsilon \quad /4/$$

where β_0 is the intercept, β_{i} is the regression coefficient associated with the i-th wavelength, X_{i} is the spectral reflectance value at the i-th channel of the AS7265x sensor, and ε is the random error term assuming a normal distribution.

Support vector regression (SVR) was implemented to handle high-dimensional non-linearity using the ϵ-insensitive loss function. The goal is to find a function f(x) that has at most ϵ deviation from the actual targets. The predicted physicochemical quality parameter using SVR can be obtained from the following equation [*18*]:

$$Y_{\text{SVR}} = f(x) = \sum_{\text{i}=1}^{18} (\alpha_{\text{i}} - \alpha_{\text{i}}^*) K(x_{\text{i}}, x) + b \quad /5/$$

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

where *K* is a kernel function, α_i and α_i^* are the Lagrange multipliers, x_i represents the support vectors, and *b* is the bias term. A Radial Basis Function (RBF) was selected as the kernel function due to its effectiveness in spectral analysis:

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \quad /6/$$

where *γ* is the kernel parameter controlling the spread of the Gaussian function, and $\|x_i\text{-}x\|^2$ is the squared Euclidean distance between the input vector and the support vectors.

An artificial neural network (ANN) based on a feed-forward multi-layer perceptron (MLP) [*19*] was designed to model the nonlinear relationship between the spectral inputs and the physicochemical quality parameters. The network consisted of an input layer with 18 neurons representing spectral responses of the AS7265x sensor, one hidden layer, and an output layer with five neurons corresponding to the estimation outputs of five physicochemical properties. The mathematical operation involves the weighted sum of inputs passed through a non-linear activation function.

For the hidden layer, the output of the j-th neuron is calculated using the hyperbolic tangent sigmoid function (tanh):

$$H_j = tanh\left(\sum_{i=1}^{18} w_{ji}^{(1)} \cdot X_i + b_j^{(1)}\right) \quad /7/$$

The final output Y_{ANN} is computed by a linear transfer function (purelin) in the output layer:

$$Y_{ANN} = \sum_{j=1}^{Nh} w_{kj}^{(2)} \cdot H_j + b_k^{(2)} \quad /8/$$

where H_j is the output of the j-th neuron, $w^{(1)}$ and $w^{(2)}$ denote the weight matrices for the hidden and output layers, respectively, $b^{(1)}$ and $b^{(2)}$ are the bias vectors, and *Nh* is the number of neurons in the hidden layer.

The number of neurons in the hidden layer is determined by using the following equation [*20*]:

$$Nh = 0.5 \cdot (Ni + No) + \sqrt{Ns} \quad /9/$$

where *Ni* and *No* are the number of input and output neurons, respectively, and *Ns* is the number of samples in the dataset.

Committee machine-based ensemble learning

To evaluate the effectiveness of ensemble integration, the estimations generated by the three base learners were combined using two distinct weighting methods, *i.e.* simple averaging (SA) and

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

genetic algorithm-based weighted averaging (GA-WA). The overall configurations of the committee machine and integration strategies are presented in Fig. 1.

(Fig. 1)

The general equation of the committee machine output (Y_{CM}) is defined as:

$$Y_{\mathrm{CM}} = \sum_{\mathrm{i}=1}^{Nb} w_{\mathrm{i}} \cdot Y_{\mathrm{i}} \qquad /10/$$

with

$$\sum_{\mathrm{i}=1}^{Nb} w_{\mathrm{i}} = 1 \qquad /11/$$

where Y_{CM} is the final prediction generated by committee machine, Y_i is the prediction produced by the i-th base learner, w_i is the corresponding weighting coefficient, and *Nb* is the total number of base learners.

The simple averaging (SA) method serves as the baseline ensemble approach. In this scheme, equal importance is assigned to all base learners, operating under the assumption that each model contributes equally to the prediction accuracy [*21*]. The weights are fixed as w_i=1/*Nb*, where in this study, *Nb* = 3, corresponding to LR, SVR, and ANN. Thus, the prediction is calculated as:

$$Y_{\mathrm{SA}} = \frac{1}{3} \times (Y_{\mathrm{LR}} + Y_{\mathrm{SVR}} + Y_{\mathrm{ANN}}) \qquad /12/$$

This method reduces the variance of the prediction error by combining the outputs of multiple learners. However, it does not consider the specific strengths or weaknesses of individual algorithms for different physicochemical parameters.

In contrast to SA, the weighted averaging (WA) approach acknowledges that base learners may exhibit varying degrees of accuracy depending on the non-linearity of the target parameter (e.g., vitamin C vs. TSS). To determine the optimal contribution of each model, a genetic algorithm (GA) was employed to optimize the weight vector w={w_{LR},w_{SVR},w_{ANN}}. The GA searched for the optimal combination of weights that minimized the prediction error of the committee machine while satisfying the constraint that the sum of the weights equals unity. The resulting weighted prediction was calculated as:

$$Y_{\mathrm{GA-WA}} = w_{\mathrm{LR}} \cdot Y_{\mathrm{LR}} + w_{\mathrm{SVR}} \cdot Y_{\mathrm{SVR}} + w_{\mathrm{ANN}} \cdot Y_{\mathrm{ANN}} \qquad /13/$$

The GA-based optimization process followed these steps:

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

1. Initialization: A population of 50 chromosomes was generated (default for number of variables ≤5), where each chromosome represented a candidate weight vector. The weights were constrained such that $\sum_{i=1}^{3} w_i = 1$ and $0 \leq w_i \leq 1$.
2. Fitness function: The fitness of each chromosome was evaluated by minimizing the mean square error (MSE) on the validation dataset:

$$\text{Fitness} = \min(MSE_{\text{val}}) \quad /14/$$

3. Evolutionary operations: To generate the next generation of weight vectors, the algorithm applied stochastic genetic operators based on MATLAB's default settings:
   a) Selection: The Stochastic Uniform selection function was employed. This method lays out a line in which each parent corresponds to a section of the line of length proportional to its scaled fitness value. The algorithm moves along the line in steps of equal size to allocate parents. This approach is preferred over simple roulette wheel selection as it preserves genetic diversity while ensuring that fitter individuals have a higher probability of being chosen for reproduction.
   b) Crossover: A crossover fraction of 0.8 was utilized with the scattered crossover function. This creates a random binary vector to select genes from the first parent where the vector is 1, and from the second parent where the vector is 0, facilitating the effective recombination of weight strategies.
   c) Mutation: A Gaussian mutation function was applied to the remaining children. This operator adds a random number derived from a Gaussian distribution (scale and shrink parameters determined by the default configuration) to the parent entries. This introduces necessary random variability to prevent the algorithm from getting trapped in local minima.
   d) Termination: The iterative process continued until the maximum number of generations (100) was reached or the average change in the fitness value was less than the function tolerance (1e-6). The final optimal weights were then applied to the test set to generate the final predictions [*22*].

Model evaluation

The dataset was randomly partitioned into a training set (80 %) and a test set (20 %). Model performance was evaluated using the coefficient of determination (R^2) and mean squared error (MSE). The coefficient of determination measures how well the regression model explains the

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

variability of the observed/actual data, while MSE measures the average squared difference between the predicted and actual values. The values of R^2 and MSE can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{Nd}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{Nd}(y_i - \bar{y})^2} \quad /15/$$

$$MSE = \frac{1}{Nd} \cdot \sum_{i=1}^{Nd}(y_i - \hat{y}_i)^2 \quad /16/$$

where y_i is actual (observed) value, $\hat{y}_i$ is predicted value, $\bar{y}$ is mean of the actual values, and Nd is the number of observations.

Statistical analysis

The physical dimensions and physicochemical properties of mango samples were statistically analysed in terms of mean value and standard deviation, as well as minimum and maximum values, using Excel [*23*] to provide information on data variability.

RESULTS AND DISCUSSION

Spectral characteristics of Arumanis mango

The average reflectance spectra of the Arumanis mango samples acquired using the AS7265x sensor are characterized by distinct absorption features across the UV-Vis-NIR range. Variations in spectral response (R(λ)) across different market sources may reflect underlying differences in post-harvest handling, ripening methods, and physiological integrity.

Spectral measurements indicate that mangoes sourced from traditional markets exhibit consistently higher spectral response across visible and near-infrared bands than those from modern markets. The spectral response curves of mango fruits obtained from modern and traditional markets, as seen in Fig. 2, exhibit a shared morphological profile across the wavelength (λ) of 410–940 nm range. This indicates that mangoes from both sources share the same basic biochemical building blocks, especially pigments and water. However, a significant quantitative divergence is observed in the magnitude of the spectral response, particularly within the spectrum of 410–700 nm.

(Fig. 2)

In spectroscopic analysis of climacteric fruits, lower reflectance (higher absorption) in the visible range is typically correlated with higher concentrations of surface pigments, such as chlorophyll-a and carotenoids [*24*]. The lower response in modern market samples suggests a denser pigment profile or a change in the optical properties of the exocarp. This phenomenon may be

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

attributed to the "stay-green" effect or delayed degradation of chlorophyll often observed in modern supply chains that utilize controlled atmosphere (CA) storage or refrigeration [*25*]. Conversely, the higher reflectance in traditional market samples may indicate a more advanced stage of chlorophyll degradation or a thinner cuticle layer due to prolonged exposure to ambient temperatures and higher transpiration rates [*26*].

Descriptive statistics of reference data

Supermarket mangoes generally offer higher and more uniform sugar levels (total soluble solids), lower surface microbial counts, and more consistent firmness and appearance. In contrast, mangoes from traditional markets show much wider variations in acidity, pH, and spoilage rates. These differences stem primarily from supply chain management, handling, and fruit selection rather than natural biological factors alone. The physical dimensions and physicochemical properties of mangoes acquired from traditional and modern markets are shown in Table 1.

(Table 1)

The observed higher and more uniform TSS in fruits from modern retail is commonly attributed to supply-chain selection and postharvest sorting: supermarkets preferentially source fruit at defined maturity stages and often apply cold-chain logistics that slow respiration and sugar loss, producing a tighter TSS distribution [*27*]. By contrast, traditional markets receive heterogeneous lots from multiple smallholders and intermediaries, increasing variance in maturity and, therefore, TSS and sugar profiles [*28*].

Acidity and pH differences follow an inductive pattern from maturity and handling: unripe or stress-affected fruits retain higher titratable acidity and lower pH, while fully ripe fruits show reduced acidity; traditional market lots therefore often show greater acidity variability because of mixed maturity and accelerated postharvest respiration under ambient conditions [*29*]. These physicochemical shifts are consistent with cultivar-level baselines but are amplified by postharvest practices [*30*].

Water content was fairly consistent between the two groups, averaging 86.4 % for modern market samples and 87.2 % for traditional ones. However, modern market samples showed a wider range of water content (81.2 to 95.2 %) compared to the narrower range in traditional markets (84.3 to 90.5 %). The maintenance of high water levels in traditional markets, despite higher ambient temperatures, may be attributed to local practices such as frequent water misting or shorter intervals between harvest and sale. Nevertheless, since average moisture levels remain nearly identical despite major differences in sugar and vitamin content, water content is a poor metric for measuring

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

market quality. Biochemical markers like sugar levels (TSS) and vitamin C (ascorbic acid) are better indicators [*31*].

Nutritionally relevant phytochemicals such as vitamin C and polyphenols show less consistent differences: when maturity and cultivar are controlled, concentrations are comparable between market types, but degradation during storage and exposure to heat/light in traditional markets can reduce ascorbic acid and antioxidant capacity, producing lower mean values and higher variance [*32,33*].

Performance comparison of base learners and committee machines

The core objective of this study was to evaluate whether the genetic algorithm (GA)-optimized committee machine could outperform single base learners. The performance metrics, represented by MSE (Table 2) and R^2 (Table 3), for the training and test sets revealed a clear hierarchy in model efficacy. Among the single base learners, linear regression (LR) yielded the lowest accuracy, particularly for vitamin C and total sugar. This underperformance is attributed to the inherent non-linearity between the spectral response and biochemical concentrations in complex fruit tissue.

(Table 2)

(Table 3)

Support vector regression (SVR) and artificial neural network (ANN) significantly outperformed LR to estimate physicochemical properties of mangoes, with ANN showing superior capability in modelling non-linearities. However, individual non-linear models occasionally suffered from overfitting, indicated by a divergence between training and test errors [*34*]. The simple averaging committee machine (SA-CM) improved stability but did not achieve the maximum potential accuracy. In contrast, the GA-optimized committee machine (GA-CM) demonstrated the highest predictive performance across all parameters.

By dynamically adjusting the weight vector w, the GA effectively assigned stronger influence to the model best suited for all parameters. The optimized weight vector w is:

$$w = \begin{bmatrix} w_{\mathrm{LR}} \\ w_{\mathrm{SVR}} \\ w_{\mathrm{ANN}} \end{bmatrix} = \begin{bmatrix} 0.0002 \\ 0.1397 \\ 0.8601 \end{bmatrix} \qquad /17/$$

This optimization resulted in a significant reduction in MSE compared to the single best model, confirming the hypothesis that ensemble techniques can compensate for low spectral resolution [*35*].

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Prediction of physicochemical attributes using GA-CM

Fig. 3a and Fig. 3b show that the GA-CM model accurately predicted total sugar content ($R^2_{\text{sugar}} = 0.986$) and TSS ($R^2_{\text{TSS}} = 0.979$). The scatter plots of measured and anticipated values showed significant linearity near the 1:1 line. This success implies that the AS7265x sensor effectively captures the C-H bond overtones in the NIR channels (810, 860, 940 nm), which are direct indicators of sweetness. The result validates the system's utility for grading mangoes based on sweetness, the primary driver of consumer preference [*36*].

Predicting vitamin C proved more challenging due to its relatively low concentration compared to sugars and water. While single models struggled, the GA-CM improved the prediction accuracy to a very good level ($R^2_{\text{vitC}} = 0.941$) as seen in Fig. 3c. This improvement suggests that the ensemble approach effectively reduces the signal-to-noise ratio inherent in low-concentration analyte detection. Similarly, pH prediction (Fig. 3d) showed reliable accuracy, likely due to its covariance with organic acids that affect the fruit's water matrix absorption properties [*37*].

As expected, water content was predicted with the highest accuracy ($R^2_{\text{water}} = 0.966$) (Fig. 3e). This is physically consistent, as water is the dominant constituent of the fruit and has a strong, broad absorption peak at 940 nm, which aligns perfectly with one of the sensor's specific channels.

(Fig. 3)

CONCLUSIONS

This study demonstrates that a genetic algorithm-optimized committee machine can significantly enhance the predictive performance of low-cost UV-Vis-NIR spectral sensors for non-destructive quality assessment of Arumanis mangoes. By integrating linear regression, support vector regression, and artificial neural networks into a weighted ensemble framework, the proposed GA-CM successfully compensated for the limited spectral resolution of the AS7265x sensor. The ensemble model achieved superior accuracy compared with single learners and simple averaging approaches across all evaluated physicochemical parameters, including total sugar content, TSS, vitamin C, pH, and water content.

The findings confirm that water content and sweetness metrics (total sugar and TSS) can be predicted with high accuracy. Furthermore, key nutritional factors like vitamin C, which are difficult to estimate using low-resolution sensors, achieved strong predictive performance when combined with ensemble optimization. Testing fruit from both modern supermarkets and traditional markets proved that the model remains dependable under diverse, real-world supply chain conditions.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Overall, this research highlights that algorithmic intelligence, rather than expensive hardware, is the key enabler for scalable and affordable non-destructive fruit quality grading. The proposed system offers strong potential for on-site application by smallholder farmers, traders, and distributors, supporting data-driven decision-making in tropical fruit supply chains. Future work should focus on expanding the dataset across seasons and cultivars, as well as integrating the system into embedded or mobile platforms for fully autonomous field deployment.

ACKNOWLEDGEMENTS

The authors express their gratitude to the Research and Community Service Institution Jenderal Soedirman University. The authors state that generative AI tools (Gemini-Edu and Grammarly) were utilized exclusively for language editing tasks, such as grammar checking, sentence enhancement, and improvement of clarity and readability of the manuscript. AI tools were not utilized to create scientific results, examine or interpret data, or draw scientific conclusions. The authors are entirely accountable for the correctness, originality, integrity, and ultimate content of the manuscript.

FUNDING

This research was funded by Research and Community Service Institution, Jenderal Soedirman University through 2025 Fundamental Research scheme under grant number No. 14.386/UN23.34/PT.01/V/2025, entitled “Development of a Portable Spectrometer Based on a Triad Spectroscopy Sensor for Non-Destructive Estimation of the Physicochemical Characteristics of Fruits Using Committee Machines and Sensory Evaluation”.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

SUPPLEMENTARY MATERIALS

Supplementary materials are available at: www.ftb.com.hr.

AUTHORS’ CONTRIBUTION

All authors fulfilled all four ICMJE authorship criteria. S.B. Sulistyo contributed mainly to conceptualisation, investigation, funding acquisition, formal analysis, and writing the original draft. D.L. Fatmala, D.Y. Putri, and M.R. Nursyahpratama were involved in data collection and analysis. K. Wijaya and P.H. Kuncoro contributed to reviewing and editing the manuscript. A. Sudarmaji was

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

responsible for resources and supervision. P. Haryanti was involved in experimental design and project administration. H.M. Kadhim corrected and critically reviewed the manuscript. All authors approved the final version of the manuscript.

ORCID ID

S.B. Sulistyo https://orcid.org/0000-0003-2892-649X
K. Wijaya https://orcid.org/0000-0003-2835-4049
P.H. Kuncoro https://orcid.org/0000-0002-1430-1598
A. Sudarmaji https://orcid.org/0000-0002-3068-7996
P. Haryanti https://orcid.org/0000-0001-5440-3452
H.M. Kadhim https://orcid.org/0000-0003-2892-649X

REFERENCES

1. Khumaidi A, Purwanto YA, Sukoco H, Wijaya SH. Using fuzzy logic to increase accuracy in mango maturity index classification: approach for developing a portable near-infrared spectroscopy device. Sensors. 2022;22(24):9704. https://doi.org/10.3390/s22249704
2. Vu ND, Nguyen VM, Tran TT. Effects of pH, total soluble solids, and pectin concentration on color, texture, vitamin C, and sensory quality of mango fruit bar. Int J Food Sci. 2023;6618300. https://doi.org/10.1155/2023/6618300
3. Zhang H, Zhan B, Pan F, Luo W. Determination of soluble solids content in oranges using visible and near infrared full transmittance hyperspectral imaging with comparative analysis of models. Postharvest Biol Tec. 2020;163,111148. https://doi.org/10.1016/j.postharvbio.2020.111148
4. Seelenbinder JA, Robb CS. Applications of portable and handheld infrared spectroscopy. In: Crocombe R, Leary P, B. Kammrath B, editors. Portable Spectroscopy and Spectrometry. Oxford, UK: John Wiley & Sons, Ltd.; 2021. https://doi.org/10.1002/9781119636489.ch33
5. Puttipipatkajorn A, Puttipipatkajorn A. Development of low-cost portable spectrometer equipped with 18-band spectral sensors using deep learning model for evaluating moisture content of rubber sheets. Smart Agric Tech. 2024;9,100562. https://doi.org/10.1016/j.atech.2024.100562

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

6. Gao J, Wang S, Erdogdu F, Marra F, Sarghini F, Chen L. Machine learning-driven innovations in food processing: A systematic review of applications, challenges, and future developments. Trends Food Sci Tech. 2025;164,105204. https://doi.org/10.1016/j.tifs.2025.105204
7. Wang C, Xu X, Zhang Y, Cao Z, Ullah I, Zhang Z and Miao M. A stacking ensemble learning model combining a crop simulation model with machine learning to improve the dry matter yield estimation of greenhouse pakchoi. Agronomy. 2024;14(8):1789. https://doi.org/10.3390/agronomy14081789
8. H Jiang, S Zhang, Z Yang, L Zhao, Y Zhou and D Zhou. Quality classification of stored wheat based on evidence reasoning rule and stacking ensemble learning. Comput Electron Agr. 2023;214;108339. https://doi.org/10.1016/j.compag.2023.108339
9. Clifford JL, Chan YJ, Yusof MAM, Tiong TJ, Lim SS, Lee CS, *et al*. Predictive modelling of H_2S removal from biogas generated from palm oil mill effluent (POME) using a biological scrubber in an industrial biogas plant: integration of artificial neural network (ANN) and process simulation. Food Technol Biotechnol. 2025;63(2):124-33. https://doi.org/10.17113/ftb.63.02.25.8792
10. Umamaheswari K, Madhumathi R. Predicting crop yield based on stacking ensemble model in machine learning. Proceedings of the 8th International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC); 2024 October 03-05; Kirtipur, Nepal; 2024. pp. 1831-6. https://doi.org/10.1109/I-SMAC61858.2024.10714785
11. Zhang P, Lu B, Shang J, Wang X, Hou Z, Jin S, *et al*. Ensemble learning for oat yield prediction using multi-growth stage UAV images. Remote Sens. 2024;16(23):4575. https://doi.org/10.3390/rs16234575
12. Steer P, Guerit L, Lague D, Crave A, Gourdon A. Size, shape and orientation matter: fast and semi-automatic measurement of grain geometries from 3D point clouds. Earth Surface Dynamics. 2022;10(6):1211 - 1232. https://doi.org/10.5194/esurf-10-1211-2022
13. Sapkota A. Anthrone test- definition, principle, procedure, result, uses. Microbe Notes; 2022. Available from: https://microbenotes.com/anthrone-test/
14. Helmenstine AM. Vitamin C Determination by Iodine Titration. ThoughtCo; 2025. Available from: https://www.thoughtco.com/vitamin-c-determination-by-iodine-titration-606322.
15. AOAC Official Method. Rockville, MD, USA: AOAC International; 2000.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

16. MATLAB, v. R2020a, The MathWorks, Inc., Natick, MA, USA; 2020. Available from: https://www.mathworks.com/products/matlab.html.
17. Laake P, Fagerland MW. Statistical inference. In: Laake P, Benestad HB, Olsen BR, editors. Research in medical and biological sciences. Amsterdam, The Netherlands: Elsevier Inc; 2015. pp. 379-430. https://doi.org/10.1016/B978-0-12-799943-2.00011-2
18. Kaloop MR, Hu JW, Elbeltagi E. Pile-raft settlements prediction under coupled static-dynamic loads using four heuristic regression approaches. Shock and Vibration. 2018; 3425461, 10 pages. https://doi.org/10.1155/2018/3425461
19. Han F, Jiang J, Ling QH, Su BY. A survey on metaheuristic optimization for random single-hidden layer feedforward neural network. Neurocomputing. 2019; 335:261–73. https://doi.org/10.1016/j.neucom.2018.07.080
20. Sulistyo SB, Sudarmaji A, Haryanti P, Kuncoro PH. A novel approach for detection of granulated coconut sugar adulteration using LED-based spectrometer and machine learning. Inf Process Agric. 2025;12(3):300-11. https://doi.org/10.1016/j.inpa.2024.09.007
21. Proskura P, Zaytsev A. Beyond simple averaging: improving NLP ensemble performance with topological-data-analysis-based weighting. Proceedings of the IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA); 2024 October 24; San Diego, CA, USA; 2024. p. 1-8. https://doi.org/10.1109/DSAA61799.2024.10722805
22. Shahhosseini M, Hu G, Pham H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. Mach Learn Appl. 2022;7;100251. https://doi.org/10.1016/j.mlwa.2022.100251
23. Microsoft Excel, v. 2024, Microsoft Corp., Redmond, WA, USA; 2024. Available from: https://www.microsoft.com.
24. Merzlyak MN, Solovchenko AE, Gitelson AA. Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. Postharvest Biol Tec. 2003;27(2):197-211. https://doi.org/10.1016/S0925-5214(02)00066-2
25. Gage E, Jain R, Terry LA, Falagán N. Controlled atmosphere as cold chain support for extending postharvest life in cabbage. Plant Physiology and Biochemistry 2024. Plant Physiol Bioch. 2024;216;109094.

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

https://doi.org/10.1016/j.plaphy.2024.109094

26. Seifert B, Pflanz M, Zude M. Spectral shift as advanced index for fruit chlorophyll breakdown. Food Bioprocess Technol. 2014;7:2050-9. https://doi.org/10.1007/s11947-013-1218-1
27. Vani G, Sivamma P, Kumar S. Physico-chemical properties of Chaunsa mangoes. Pharma Innov J. 2021;10(4):1127-31. Available from: https://www.thepharmajournal.com/archives/?year=2021&vol=10&issue=4&ArticleId=6155
28. Partibha, Gehlot R, Rekha, Kumari A, Sindhu R, Singh R, *et al*. Physico-chemical characteristics of mature green mango fruit and mint leaves. Int J Chem Stud 2020;8(6):2607-2609. https://doi.org/10.22271/chemi.2020.v8.i6ak.11177
29. Dhumal AD, Swami SB. Physico-chemical properties of alphonso mango pulp. Int J Food Ferment Technol. 2022;12(1):19-24. https://doi.org/10.30954/2277-9396.01.2022.3
30. Kuldeep DK, Sharma DP, Bhooriya MS, Nagwanshi A, Kumar M, Uikey V, *et al*. A comprehensive review on physicochemical analysis and quality parameters of various mango fruit cultivars. Biol Forum Int J. 2023;15(10):1384-9. Available from: https://www.researchtrend.net/bfij/current_issue_bfij.php?taxonomy-id=168
31. Simkova K, Veberic R, Hudina M, Grohar MC, Pelacci M, Smrke T, *et al*. Non-destructive and destructive physical measurements as indicators of sugar and organic acid contents in strawberry fruit during ripening. Sci Hortic-Amsterdam. 2024;327;112843. https://doi.org/10.1016/j.scienta.2024.112843
32. Sabuz AA, Molla MM, Khan MHH, Chowdhury MGF, Pervin S, Alam M, *et al*. Comparison of physicochemical composition, antioxidant activity and common phytochemicals of selected BARI mango varieties and commercial cultivar, Langra. Food Chem Adv. 2024;4;100580. https://doi.org/10.1016/j.focha.2023.100580
33. Chandel R, Sharma PC. Physico-chemical properties and arsenic residue patterns in locally ripened mango fruits. Int J Hortic Agric Food Sci. 2025;9(1):7-16. http://dx.doi.org/10.22161/ijhaf.9.1.2
34. Wang C, Xu X, Zhang Y, Cao Z, Ullah I, Zhang Z, *et al*. A stacking ensemble learning model combining a crop simulation model with machine learning to improve the dry matter yield estimation of greenhouse pakchoi. Agronomy. 2024;14(8):1789. https://doi.org/10.3390/agronomy14081789

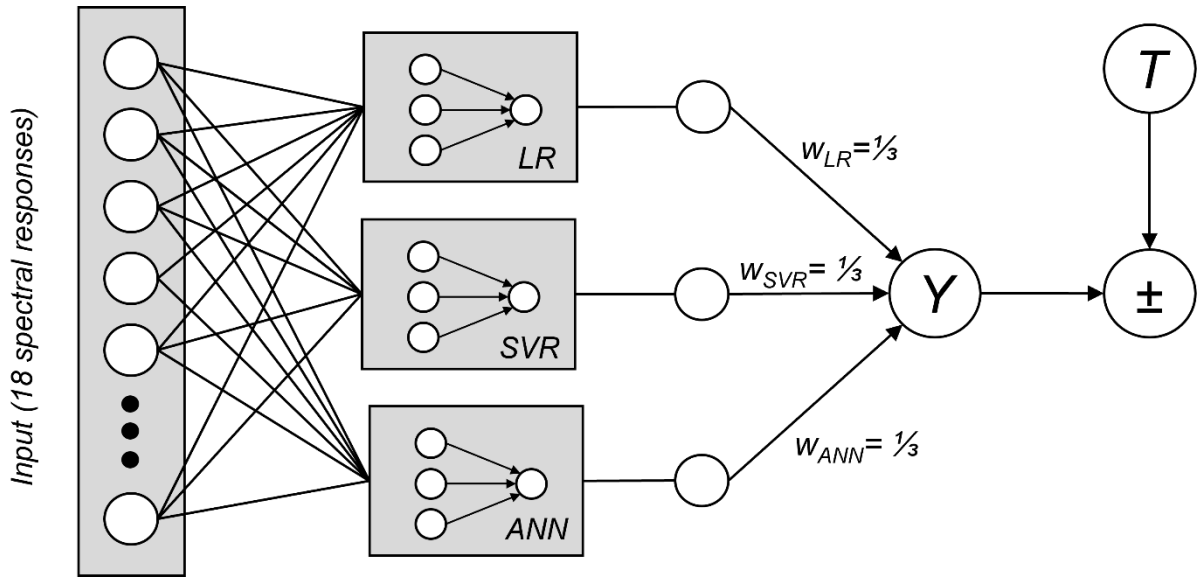
Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

35. Wang K, Bian X, Tan X, Wanga H, Li Y. A new ensemble modeling method for multivariate calibration of near infrared spectra. Anal Methods. 2021;13:1374-80. https://doi.org/10.1039/D1AY00017A

36. Nguyen C, Phan Q, Tran N, Fukuzawa M, Nguyen P, Nguyen C. Precise sweetness grading of mangoes (*Mangifera indica* L.) based on random forest technique with low-cost multispectral sensors. IEEE Access. 2020;8: 212371-82. https://doi.org/10.1109/ACCESS.2020.3040062

37. Li J, Huang W, Zhao C, Zhang B. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. J Food Eng. 2013;116(2):324-332. https://doi.org/10.1016/j.jfoodeng.2012.11.007

a)



Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

b)

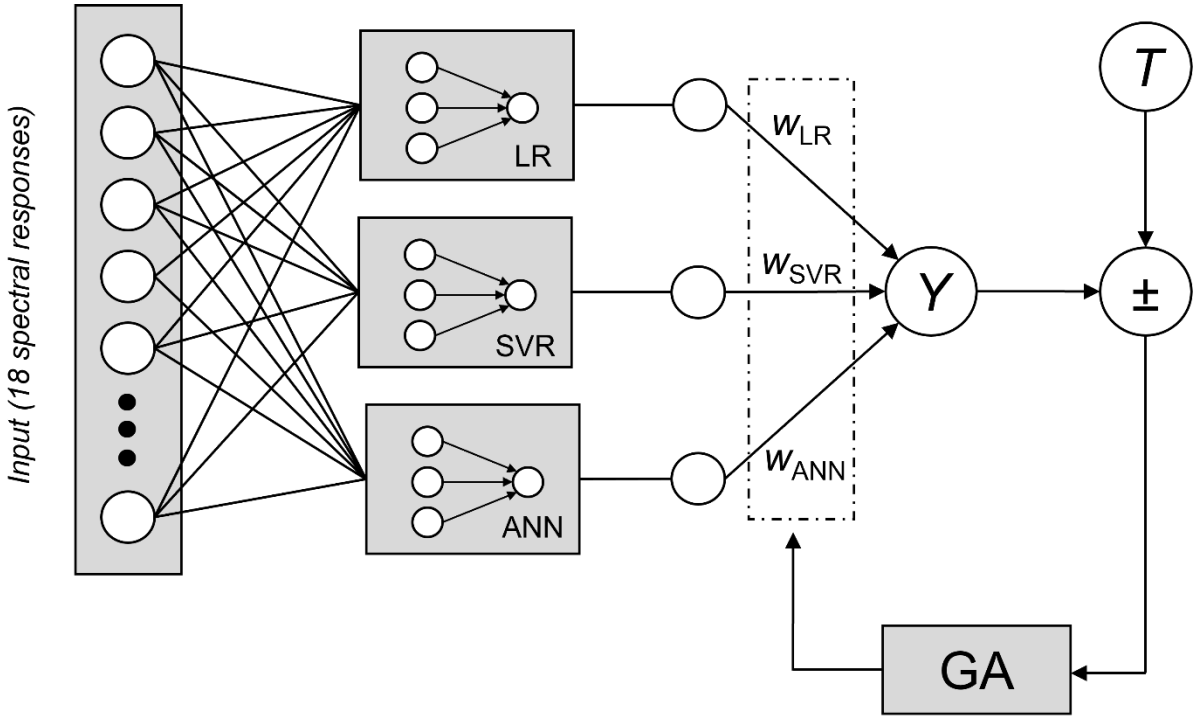
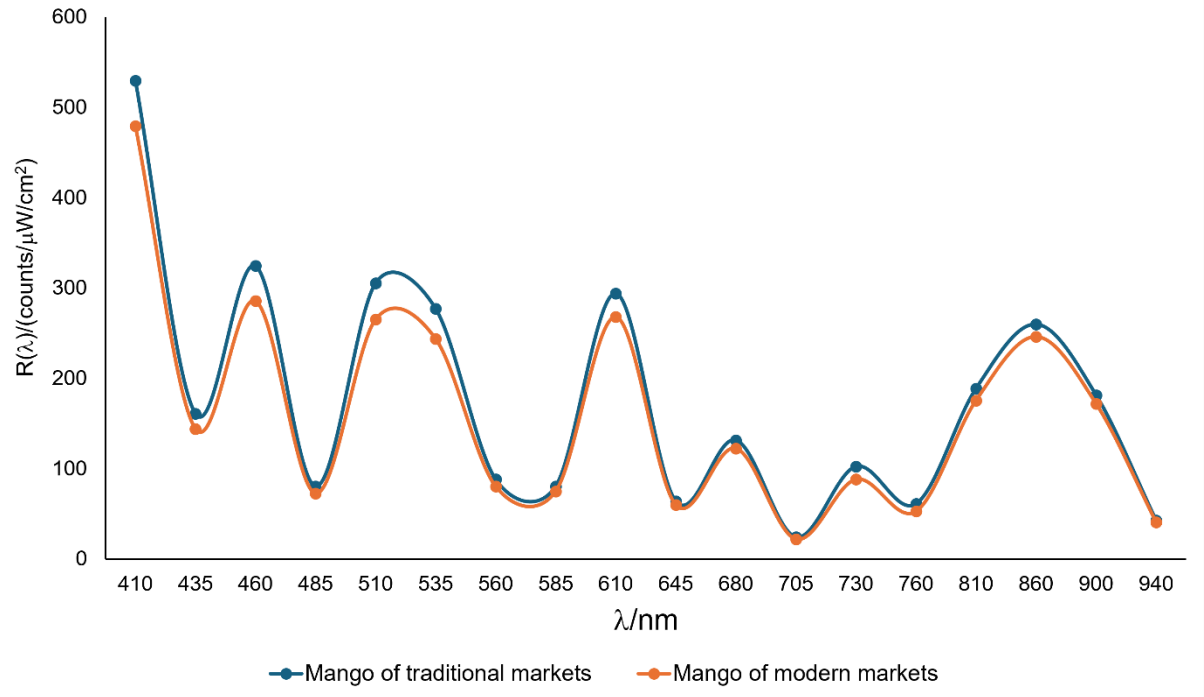


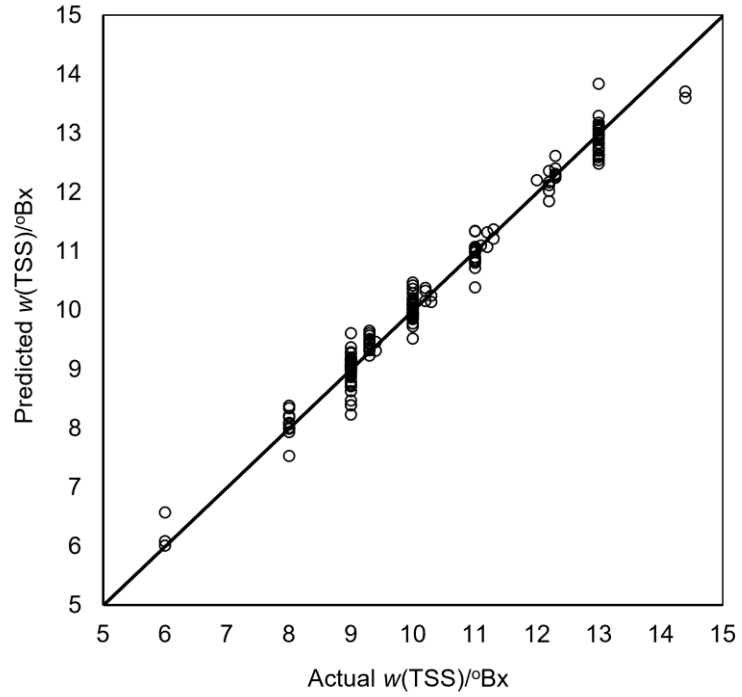
Fig. 1. The architecture of the developed committee machines: a) simple averaging, b) GA-based weighted averaging



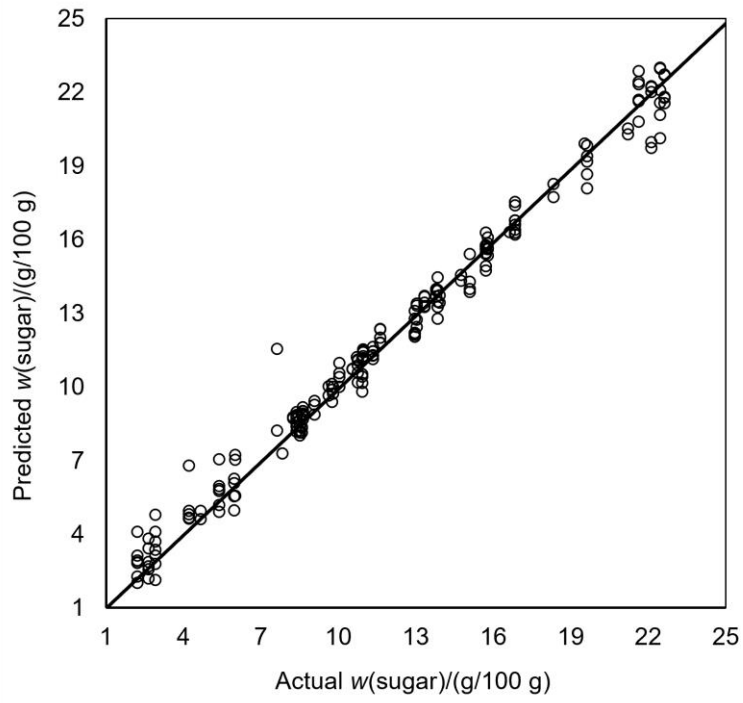
Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Fig. 2. The variations in spectral response of mangoes acquired from different market types

a)

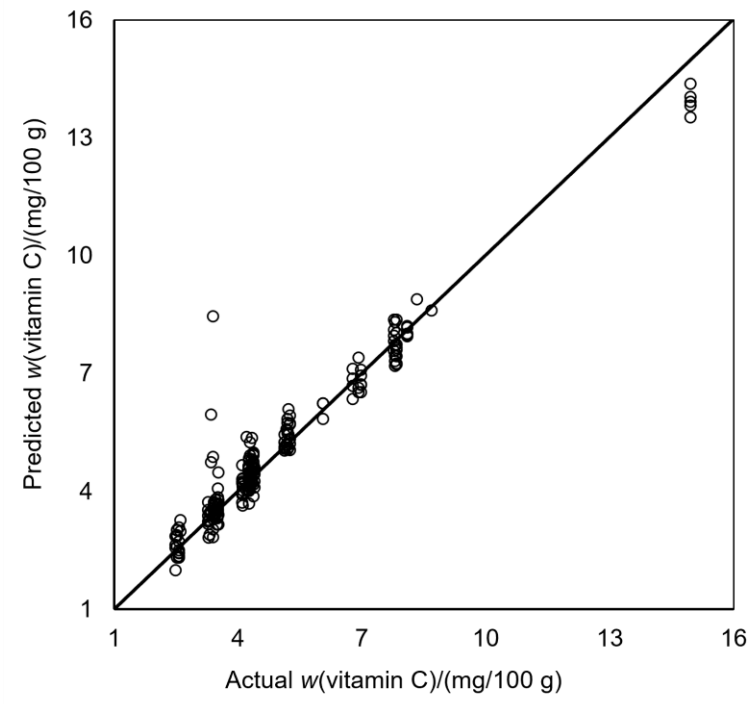


b)

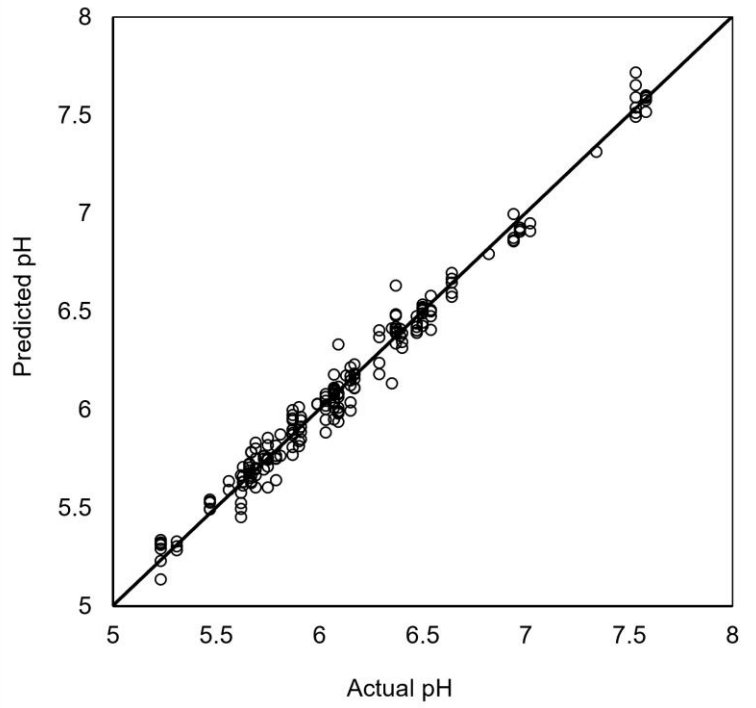


Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

c)



d)



Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

e)

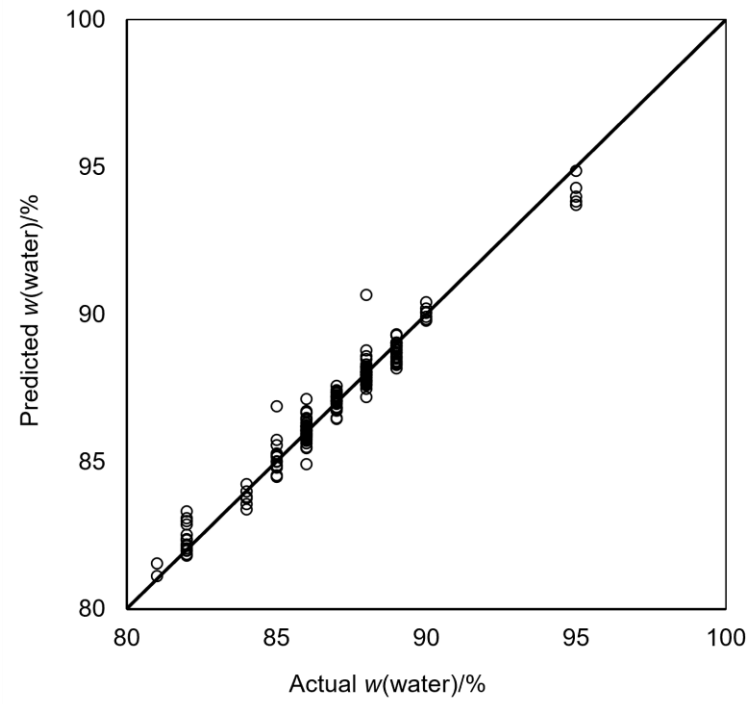


Fig. 3. The scatter plots of actual and predicted values of: a) TSS, b) sugar content, c) pH, d) vitamin C, and e) water content of mango

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

Table 1. Physical dimensions and physicochemical properties of mangoes acquired from traditional and modern markets

Parameter	Traditional markets	Modern markets
m/g	(325.3±39.1)[a]	(330.8±45.2)[a]
L/mm	(133.4±6.2)[a]	(132.7±5.6)[a]
W/mm	(82.4±3.4)[a]	(82.7±3.8)[a]
T/mm	(68.5±3.1)[a]	(69.2±3.3)[a]
A/cm²	(275.6±20.0)[a]	(276.8±20.0)[a]
w(vitamin C)/(mg/100 g)	(4.31±1.62)[a]	(5.52±2.55)[b]
w(water)/%	(87.2±1.4)[a]	(86.4±3.2)[a]
TSS/°Bx	(9.40±1.27)[a]	(10.85±1.48)[b]
pH	(6.06±0.50)[a]	(6.17±0.62)[a]
w(total sugar)/(g/100 g)	(9.43±5.13)[a]	(14.46±4.20)[b]

L=fruit length, *W*=fruit width, *T*=fruit thickness, *A*=fruit surface area, TSS=total soluble solids. Mean values followed by the same lowercase letter in the row do not differ significantly from each other ($p>0.05$) according to Tukey's test

Table 2. Comparison of MSE results from base learners and committee machines

Physicochemical property	LR	SVR	ANN	SA-CM	GA-CM
w(vitamin C)/5	4.24680	1.04325	0.36629	1.05854	0.31990
w(water)/%	0.043	0.011	0.002	0.010	0.002
TSS/°Bx	1.61734	0.39278	0.06133	0.36074	0.05327
pH	0.20263	0.02682	0.00891	0.03659	0.00761
w(total sugar)/%	22.99595	3.75156	0.48720	4.70261	0.50759
Average MSE	5.81263	1.04290	0.18475	1.23172	0.17768

Table 3. Comparison of R^2 results from base learners and committee machines

Physicochemical property	LR	SVR	ANN	SA-CM	GA-CM
w(vitamin C)/%	0.1817	0.8647	0.9336	0.9096	0.9409
w(water)/%	0.2304	0.8657	0.9614	0.9095	0.9660
TSS/°Bx	0.3794	0.8855	0.9769	0.9292	0.9795
pH	0.3278	0.9273	0.9711	0.9299	0.9748
w(total sugar)/%	0.3039	0.9211	0.9854	0.9333	0.9862
Average R^2	0.2846	0.8929	0.9657	0.9223	0.9695

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

SUPPLEMENTARY MATERIAL

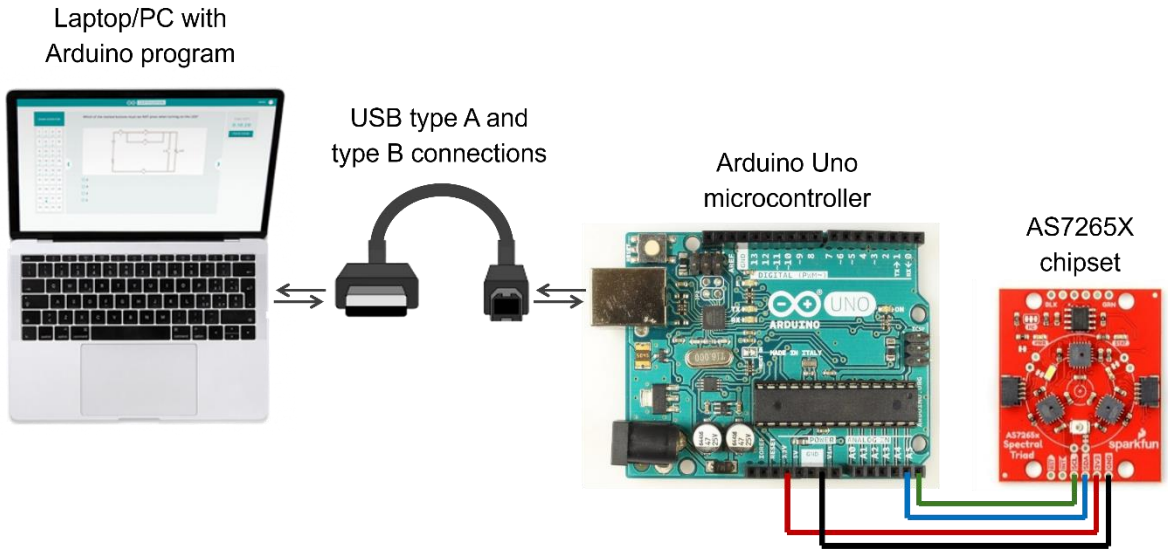


Fig. S1. The developed prototype integrating the AS7265x chipset coupled with Arduino Uno microcontroller

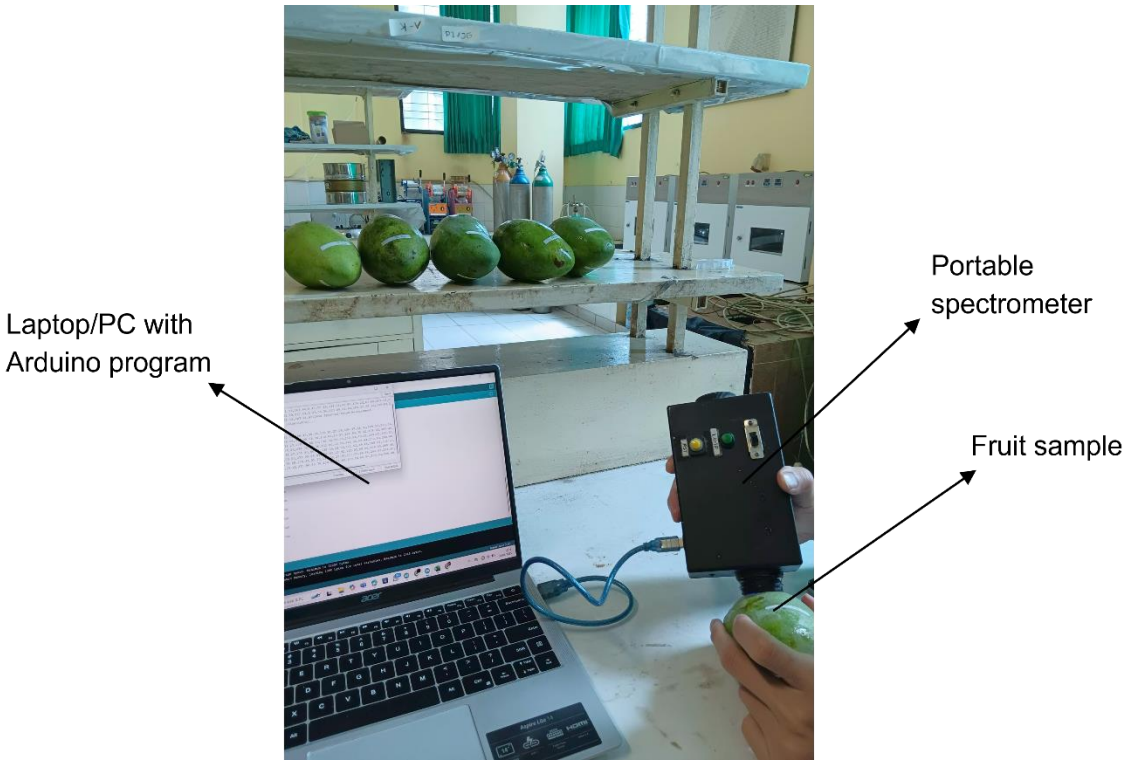


Fig. S2. Acquisition of mango spectral responses using the developed portable spectrometer

Please note that this is an unedited version of the manuscript that has been accepted for publication. This version will undergo copyediting and typesetting before its final form for publication. We are providing this version as a service to our readers. The published version will differ from this one as a result of linguistic and technical corrections and layout editing.

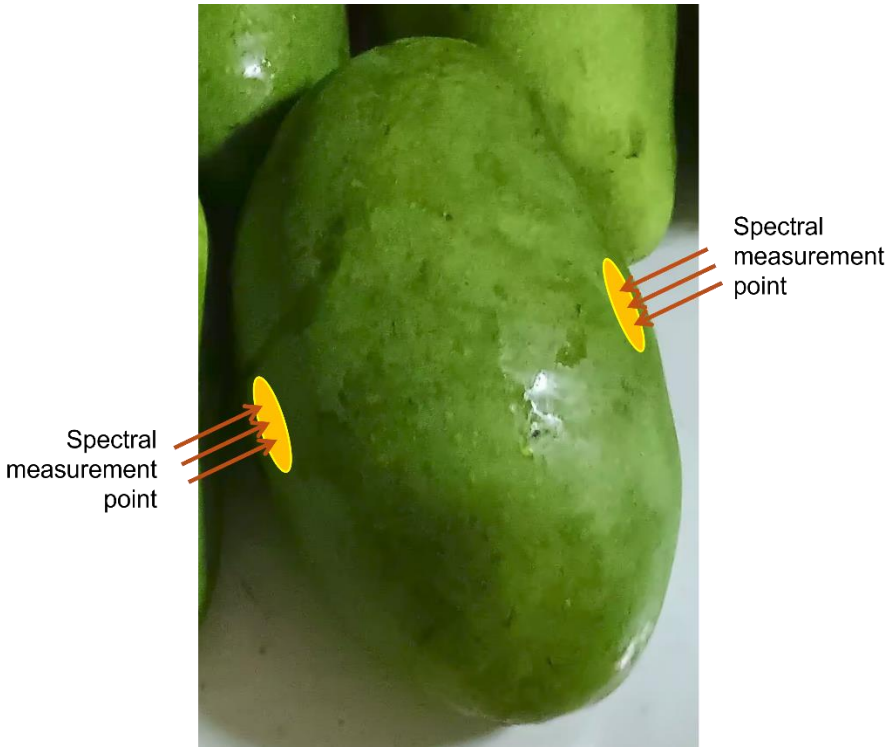


Fig. S3. Measurement points of spectral responses on each mango sample

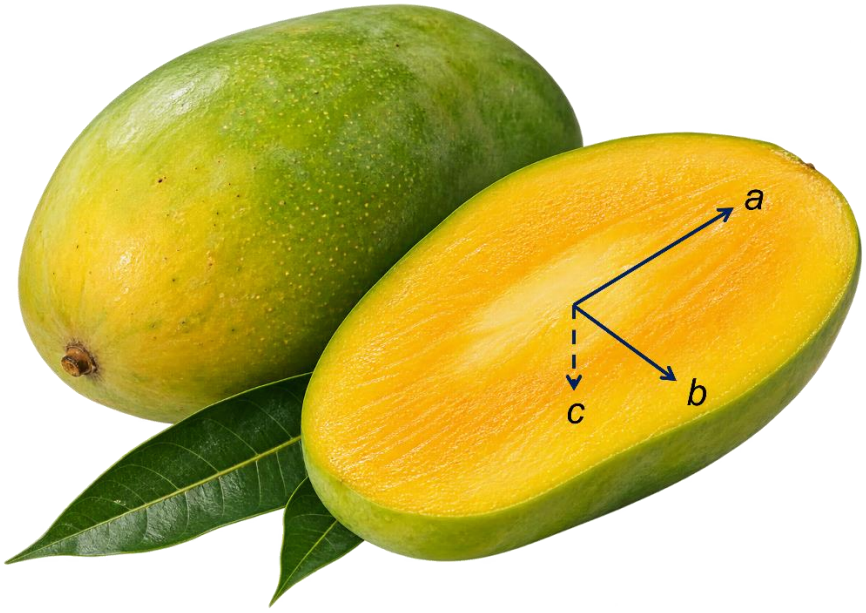


Fig. S4. Length (*a*), width (*b*), and thick (*c*) axis of mango for measuring fruit size and surface area