Automated Interactance Near-Infrared Spectral Acquisition System for Mandarin Quality Assessment

Van-Linh Lam[1], Dinh-Tri Nguyen[2], Thanh-Trung Le[3], Hoang-Tien Nguyen[4],
Phuoc-Loc Nguyen[5], Quoc-Khanh Huynh[6], Nhut-Thanh Tran[7], Chanh-Nghiem Nguyen[8]*
Faculty of Automation Engineering-College of Engineering, Can Tho University, Can Tho, Vietnam[1,2,3,4,7,8]
Automation Laboratory-College of Engineering, Can Tho University, Can Tho, Vietnam[1,2,3,4,8]
Faculty of Electric-Electronics, Kien Giang College, An Giang, Vietnam[1]
Faculty of Electric-Electronics and Computer, Kien Giang Vocational College, An Giang, Vietnam[5]
Faculty of Mechanical Engineering-College of Engineering, Can Tho University, Can Tho, Vietnam[6]

***Abstract*—Intelligent fruit grading systems require automated sensing solutions capable of rapid and reliable non-destructive internal quality assessment. While near-infrared spectroscopy has been widely used for soluble solids content (SSC) prediction, most existing studies rely on manually acquired spectra, limiting scalability in smart agricultural environments. Although online Vis–NIR systems based on transmission configurations have been reported, automated interactance-based systems designed for deployment-oriented grading remain limited. This study presents the design and validation of an automated interactance near-infrared spectral acquisition system for mandarin SSC evaluation. The system integrates controlled clamping, rotational positioning, and automated probe actuation to ensure stable optical geometry and repeatable probe–fruit contact during measurement. Spectral consistency was assessed by comparing consecutive scans obtained using manual and automated acquisition modes. The automated system reduced spectral dispersion among consecutive acquisitions within a measurement session by approximately 65% relative to manual measurement, indicating improved acquisition stability. Chemometric models based on partial least squares regression, support vector regression, and extremely randomized trees were developed under multiple preprocessing strategies. Prediction performance under automated acquisition remained within the same range as manual measurement, with several preprocessing–model combinations (particularly PLS and SVR with smoothing-based preprocessing), showing slightly higher R_{p} values and lower $\mathrm{RMSE}_{\mathrm{p}}$ values under automated acquisition. The findings demonstrate the feasibility of the proposed automated system for stable interactance spectral acquisition suitable for SSC prediction, supporting its potential future integration into automated fruit quality assessment systems.**

***Keywords*—Automated spectral acquisition; interactance spectroscopy; near-infrared spectroscopy; fruit quality assessment; soluble solids content; chemometric modeling; smart agriculture; sensing systems**

I. Introduction

Fruit quality evaluation plays a critical role in postharvest handling, grading, and commercialization of horticultural products. With the rapid development of smart agricultural technologies, there is increasing demand for automated sensing systems capable of delivering rapid, nondestructive, and reliable internal quality assessment under practical operational conditions. Among quality attributes, sweetness, commonly quantified by soluble solids content (SSC), is one of the most important indicators for mandarins. Conventional SSC measurement using refractometers requires juice extraction, making it destructive, time-consuming, and unsuitable for integration into automated grading systems [1]. Consequently, nondestructive optical sensing techniques have attracted increasing attention [2], particularly near-infrared (NIR) spectroscopy.

In NIR spectroscopy, incident radiation interacts with organic constituents within the fruit, inducing overtone and combination vibrations of C–H, N–H, and O–H bonds. The resulting spectral signatures contain information related to key quality attributes and can be effectively exploited using chemometric and machine-learning models for fruit quality prediction [3]. Depending on the optical configuration, spectral data can be acquired in reflectance, transmittance, or interactance modes. While reflectance measurements are influenced primarily by surface properties, transmittance measurements require high-intensity light sources to penetrate the entire sample. Interactance mode represents a practical compromise, capturing diffusely transmitted light that has penetrated into the fruit tissue, thereby enhancing sensitivity to internal quality attributes [4], [5].

Recent studies have developed online Vis–NIR systems for fruit quality assessment, mainly using transmission or reflectance configurations integrated with conveyor-based transport mechanisms [6], [7], [8], [9], [10]. For example, Wang et al. [10] reported an online diffuse transmission system for SSC detection in pomelo. Although such transmission systems enable continuous measurement, they usually require high-intensity light sources and are sensitive to fruit size, peel thickness, and internal structure. In contrast, interactance mode measures light that has penetrated into the fruit under controlled probe–fruit contact conditions, which can improve sensitivity to internal attributes while maintaining moderate optical penetration depth. However, automated interactance acquisition systems specifically designed to ensure controlled mechanical positioning and stable optical geometry for deployment-oriented grading applications remain relatively

*Corresponding author

limited.

For citrus fruits, peel composition and thickness significantly influence optical measurements. Although citrus peels may contain small amounts of free sugars, their contribution is relatively limited compared with the juice sac content [11], [12], suggesting that surface-dominated measurements may inadequately reflect internal sweetness. Interactance spectroscopy is therefore particularly suitable for mandarin SSC evaluation, as the porous and moderately thin peel permits partial penetration of NIR radiation into juice-containing tissues.

Despite extensive research on NIR-based fruit quality prediction, many reported systems remain laboratory-oriented and rely on manually operated spectral acquisition procedures. However, manual probe positioning can introduce variability associated with contact pressure, alignment differences, and operator handling, thereby limiting repeatability and scalability in automated grading environments. For practical smart agricultural deployment, sensing systems must ensure consistent optical conditions during measurement.

Preliminary feasibility experiments demonstrated that automated interactance spectral acquisition reduced the average measurement time per fruit from approximately 40 s to 17.7 s, corresponding to a 55.75% reduction and an estimated throughput of approximately 203 fruits per hour, while also improving measurement stability. At the same time, SSC prediction performance remained comparable to that of manual measurement, even under limited dataset size, preprocessing strategies, and regression models. These findings indicate that automation does not inherently degrade chemometric modeling performance and provide a foundation for further system-level development. However, these preliminary results remain limited to feasibility-level validation and do not yet fully address deployment-level requirements, including robustness, scalability, and operational reliability, for practical quality assessment systems. In addition, the study remains limited in terms of modeling configurations.

To address this gap, an automated interactance NIR spectral acquisition system is proposed. The system is an integrated sensing framework in which a dedicated acquisition module functions as the core measurement unit, coupled with data processing and modeling components for SSC prediction, thereby enabling integration into intelligent mandarin quality assessment systems. Within this framework, the acquisition module integrates controlled clamping, rotational positioning, and automated probe actuation to enable standardized, repeatable, and geometry-consistent spectral acquisition.

The proposed system emphasizes automated fruit positioning and controlled optical geometry to ensure repeatable spectral acquisition under operational conditions, thereby supporting reliable deployment in practical grading environments. Although the system was evaluated using mandarins in this study, the sensing architecture and acquisition strategy can be adapted to fruits with similar peel thickness and internal tissue structure, offering a scalable foundation for spectroscopy-based automated grading systems.

To evaluate whether automated acquisition preserves chemometric modeling performance, regression models including partial least squares regression, support vector regression, and extremely randomized trees were developed under multiple preprocessing strategies using a consistent cross-validation framework. Comparative analysis between manual and automated modes was conducted to determine whether observed differences arise from sensing configuration rather than modeling characteristics.

II. Methodology

A. Dataset Preparation

Fig. 1 illustrates the manual interactance spectral acquisition system. A halogen light source (MI-152, Dolan-Jenner) with a lamp power of 150 W was employed with sufficient intensity to penetrate the mandarin fruit, thereby satisfying the requirements for interactance spectral acquisition. Light from the source was delivered to the fruit through an MA4 ring-light optical fiber supplied by the manufacturer. The fiber output formed a circular illumination pattern that ensured uniform illumination at the measurement location. A liquid light guide (LLG5-4Z, Thorlabs) was used to transmit the interactance signal to the spectrometer (DLP NIRscan Nano EVM, Texas Instruments) which is capable of acquiring spectra in the wavelength range of 900–1700 nm. However, preliminary inspection indicated that signal intensities beyond approximately 1365 nm were extremely low ($< 3\%$ for automated acquisition and $\approx 0.6\%$ for manual acquisition), suggesting a noise-dominated region. Therefore, only the spectral range of 900–1365 nm was retained for subsequent analysis.

To enlarge the dataset, interactance spectra were collected from four measurement regions uniformly distributed along the equator of each mandarin. For each region, three adjacent measurement points were acquired and averaged to produce one representative spectrum used for modeling. A total of 460 mandarin fruits were used in this study. Because one representative spectrum was obtained from each of four measurement regions per fruit, the full dataset comprised 1840 averaged spectra for each acquisition mode (manual and automated). The spectral acquisition procedure for each region is illustrated in Fig. 2. Three measurement points were defined along the equatorial line of the fruit (Fig. 2a). A central point was first selected on the equator, followed by two additional points positioned 1 cm to the left and right of the central location. Spectra collected at the three measurement points were averaged to obtain a representative spectrum for the measurement region. This averaging strategy was adopted to reduce measurement noise and to mitigate the influence of spatial heterogeneity in SSC distribution.

B. Manual Acquisition of Interactance Spectra

At each measurement point, the mandarin was maintained in direct contact with the probe head (Fig. 2b) to block ambient light and ensure that only interactance light emerging from within the fruit was captured. This manual interactance spectral acquisition procedure is relatively time-consuming and is therefore mainly suitable for feasibility studies and laboratory-scale investigations. To overcome these limitations and enable high-throughput, non-destructive mandarin quality assessment, an automated interactance spectral acquisition module was developed, as described in the following section.

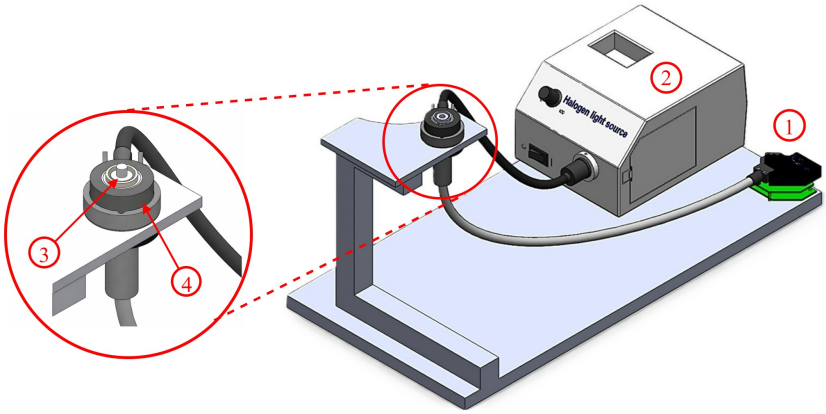
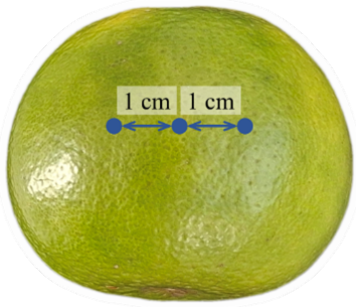
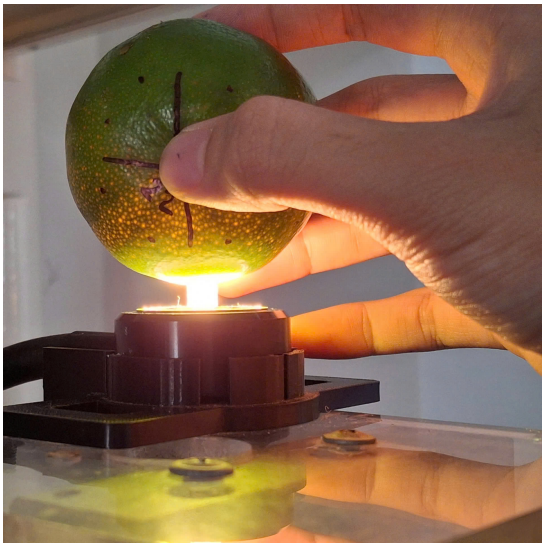


Fig. 1. Setup for collecting interactance spectra manually, including: (1) DLP NIRScan Nano EVM sensor, (2) MI-152 halogen light source, (3) the probe head, and (4) ring light output.



(a)



(b)

Fig. 2. (a) Spectral acquisition positions; (b) Interactance spectral measurement using the manual module.

C. Automated Interactance Spectral Acquisition

1) Overview of the system: The automated interactance spectral acquisition system provides controlled mechanical positioning and stable optical conditions for repeatable spectral measurements of mandarins. The system integrates fruit positioning, probe actuation, and optical isolation within a compact structure to enable automated multi-location spectral acquisition.

The overall mechanical configuration is shown in Fig. 3. The system is housed in a rigid frame forming a dark chamber that minimizes ambient light interference during measurement. Inside the chamber, a clamping and rotation mechanism positions the fruit, while an actuated spectral probe performs interactance measurements at predefined locations.

The clamping mechanism secures the mandarin between two curved contact heads to maintain stable positioning during measurement. Fruit presence and clamp position are monitored by sensors and limit switches to ensure consistent contact without excessive mechanical loading. The clamping frame is mounted on a rotary stage driven by a stepper motor, enabling controlled angular positioning for multi-location spectral acquisition.

The spectral probe–illumination unit is mounted on a horizontally actuated linear mechanism. During operation, the probe is automatically moved toward the fruit surface until a limit switch indicates contact, ensuring a repeatable probe–fruit interface. The probe integrates a ring-shaped illumination source surrounding the optical fiber used for spectral acquisition, enabling interactance measurements while reducing specular reflection.

To maintain measurement stability, the probe position, fruit orientation, and measurement sequence are automatically controlled by the system controller. This configuration enables repeatable spectral acquisition at multiple positions on the fruit surface while minimizing operator intervention.

2) Workflow of the system: Fig. 4 summarizes the operating workflow of the automated interactance spectral acquisition system. After fruit detection and clamping, the probe–illumination unit is advanced to acquire an interactance spectrum at a predefined measurement point, defined by a controlled angular position of the fruit. The fruit is then rotated sequentially to subsequent measurement points, and spectra are acquired following the same procedure. In this

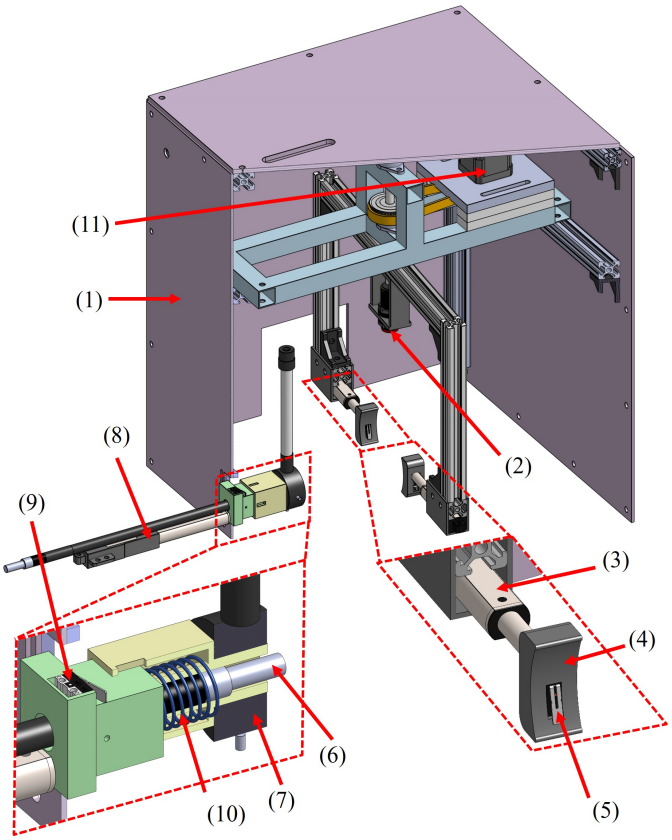


Fig. 3. Automated interactance spectral acquisition system and detailed views of its key mechanical components: (1) Dark chamber frame; (2) Proximity sensor for mandarin detection; (3) Clamp actuator; (4) Curved fruit contact head; (5) Integrated limit switch for contact detection; the acquisition module consisting of (6) Spectral probe head, (7) Ring-shaped illumination source, (8) Spectral probe–illumination unit actuator, (9) Limit switch for probe–fruit contact detection, (10) Compression spring for compliant probe contact; and (11) Stepper motor for rotating the clamp.

study, interactance spectra are collected at three circumferential measurement points per fruit to improve the spatial representativeness of the spectral measurement. Upon completion of the acquisition sequence, the probe retracts, the clamps release the fruit, and the probe returns to its initial state, ready for the next measurement cycle.

The automated interactance spectral acquisition system does not require manual predefinition of measurement positions. Based on a preliminary survey of the experimental samples, the average diameter of the sweet mandarins used for spectral acquisition was approximately 6.7 cm. Let $l = 1$ cm denote the arc length between two adjacent measurement points, $R \approx 3.35$ cm the average mandarin radius, and α (°) the required rotation angle of the fruit center to achieve an arc length of l. The rotation angle between two adjacent measurement points is calculated as:

$$\alpha = \frac{180\,l}{R\pi} = \frac{180 \times 1}{3.35\pi} \approx 17.1^\circ \tag{1}$$

The clamping frame is rotated via a belt transmission with a gear ratio $i = 1/2$ and driven by a stepper motor with a resolution of $N = 800$ steps per revolution. To rotate the

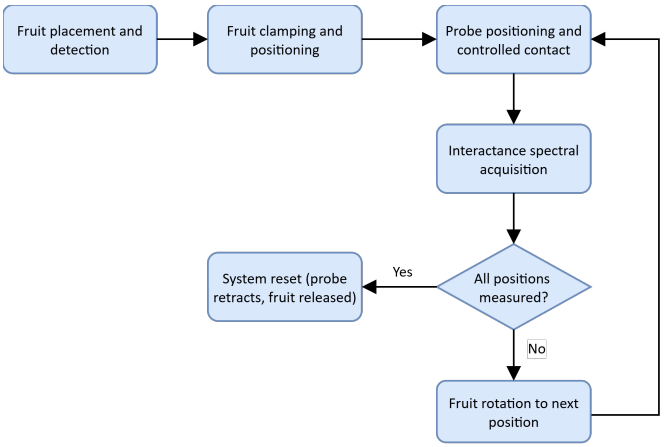


Fig. 4. Workflow of the automated interactance spectral acquisition system.

mandarin by an angle α, the required number of motor steps n is given by:

$$n = \frac{N\,\alpha}{360\,i} = \frac{800 \times 17.1}{360 \times 2} \approx 19 \text{ steps} \tag{2}$$

D. Performance Evaluation

1) Preliminary assessment of spectral consistency: To verify that the automated interactance spectral acquisition system provides spectra comparable in acquisition stability to those obtained using the manual setup, a preliminary assessment was conducted. This analysis focuses on evaluating the spectral consistency of consecutive acquisitions performed within a single measurement session, providing an initial indication of acquisition stability prior to regression model development. Improved spectral consistency is expected to enhance the reliability of the acquired data and is therefore likely to be reflected in the performance of subsequent chemometric models.

For each acquisition mode (manual and automated), four interactance spectra were acquired from consecutive acquisitions at the same nominal measurement region on the fruit surface under identical operating conditions within a single measurement session. Between consecutive acquisitions, the probe was repositioned according to the standard operating procedure of each mode. Consequently, measurements were performed within the same region but not at an identical physical contact point, thereby reflecting practical measurement variability associated with probe handling and positioning.

To quantify the consistency among repeated spectral measurements, the root mean square deviation (RMSD) was calculated as:

$$\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2} \tag{3}$$

where, n denotes the number of wavelengths in the spectrum and d_i is the standard deviation of the four repeated measurements at wavelength i, defined as:

$$d_i = \sqrt{\frac{1}{3}\sum_{j=1}^{4}(y_{ij} - \bar{y}_i)^2} \quad (4)$$

where, y_{ij} represents the interactance amplitude of the j-th replicate at wavelength i, and $\bar{y}_i$ is the mean amplitude at wavelength i calculated from the four repeated measurements. RMSD provides a wavelength-averaged indicator of spectral dispersion across the full spectral range and reflects variability among consecutive acquisitions.

A smaller RMSD indicates higher spectral consistency and, therefore, greater acquisition stability under the specified conditions. Because the measurements were conducted within a single session and confined to the same fruit region, this analysis primarily reflects instrumental and operational stability rather than broader sources of variability. For simplicity, inter-session variability across different days, operators, or fruit batches was not evaluated.

Therefore, this comparison should be interpreted as a screening-level assessment of acquisition stability. The overall feasibility of the automated acquisition system was further evaluated based on regression models calibrated using spectra acquired under each mode. Demonstrating reduced spectral dispersion under automated operation supports the reliability of the acquired data and confirms that improved spectral consistency does not adversely affect regression performance compared with manual acquisition. This result indicates that the automated acquisition system can provide spectral data of sufficient stability for subsequent regression model development for SSC prediction.

2) Regression model development: Regression models were developed to predict mandarin SSC (°Brix) from Vis–NIR interactance spectra acquired under both manual and automated acquisition modes. The primary objective was to evaluate whether spectra obtained using the automated interactance spectral acquisition system can support predictive models with performance comparable to those developed from manually acquired spectra.

As described in Section II-A, the study used 460 mandarin fruits, each measured at four equatorial regions, yielding 1840 averaged spectra per acquisition mode. The corresponding reference SSC values of these fruits ranged from 8.0 to 13.9 °Brix (mean 10.3 ± 1.14). For regression modeling, the dataset for each acquisition mode was partitioned into calibration (80%) and prediction (20%) subsets using the Sample set Partitioning based on joint X–Y distances (SPXY) algorithm [13], which selects representative samples by maximizing joint distances in both spectral (X) and response (SSC, Y) spaces. This strategy has been widely adopted to improve calibration representativeness and prediction reliability in NIR-based fruit quality modeling [14], [15], [16].

Prior to model calibration, spectral preprocessing was applied to reduce variability associated with scattering effects, baseline drift, and instrumental noise. The evaluated preprocessing methods included Min–Max scaling (MM), Standard Normal Variate (SNV), Savitzky–Golay (SG) smoothing, first-derivative (FD) transformation, and Mean–Median Absolute Deviation Normalization (MMADN), representing scaling, scatter correction, smoothing, derivative enhancement, and robust normalization strategies, respectively [17].

TABLE I. HYPERPARAMETER SEARCH RANGES AND FIXED SETTINGS FOR PLSR, SVR, AND ETR MODELS OPTIMIZED USING 10-FOLD CROSS-VALIDATION FOR MANDARIN SSC (BRIX) PREDICTION.

Model	Hyperparameter	Search range / fixed value
PLS	n_components	1 – N (number of wavelengths)
SVR	C	5 – 300 (step size: 5)
	ϵ (epsilon)	0.05–1.00 (step size: 0.05)
	cache_size	50, 100, 150
	Kernel	RBF
	Gamma	scale
ETR	n_estimators	10 – 200 (step = 10)
	max_depth	2 – 10 (step = 2)
	min_samples_split	default (2)
	min_samples_leaf	default (1)
	max_features	default (auto)
	bootstrap	False
	random_state	42
	criterion	squared_error

Because preprocessing selection can substantially influence model behavior and address different types of spectral artefacts [18], both single and combined preprocessing strategies were systematically evaluated to ensure a fair and unbiased comparison between manual and automated acquisition modes. Identical preprocessing procedures were applied to spectra from both acquisition modes to ensure comparability of regression results.

Three regression algorithms were investigated: Partial Least Squares regression (PLSR), Support Vector Regression (SVR), and Extremely Randomized Trees or Extra-Trees (ETR). PLSR was adopted as a baseline chemometric model, as it is described as a conventional regression method for SSC prediction in recent NIR fruit quality studies [19]. SVR was included to account for potential nonlinear relationships between spectral features and SSC, consistent with recent fruit spectroscopy studies that apply support vector machine–based models alongside PLSR [2]. ETR was incorporated as an ensemble tree-based regression method. Ensemble learning approaches have been reported as increasingly applied in near-infrared spectroscopy combined with chemometrics methods [20].

Model hyperparameters were optimized using grid search combined with 10-fold cross-validation performed exclusively on the calibration subset. The same validation protocol, search ranges (Table I), and evaluation criteria were applied across preprocessing–model combinations and for both acquisition modes to maintain methodological consistency.

3) Model performance evaluation metrics: The predictive performance of regression models was evaluated using four commonly reported metrics: the correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), and residual prediction deviation (RPD). These metrics are widely adopted in NIR spectroscopy studies for fruit quality prediction and provide complementary perspectives on model accuracy [9], [10], [21].

The correlation coefficient R reflects the linear relation-

ship between reference and predicted SSC values, while the coefficient of determination R^2 indicates how much of the variability in the reference measurements is explained by the model. RMSE was calculated for both calibration ($\mathrm{RMSE_c}$) and prediction ($\mathrm{RMSE_p}$) sets as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

where, y_i and $\hat{y}_i$ denote reference and predicted SSC values, respectively, and n is the number of samples. RMSE provides a scale-dependent measure of prediction error and is commonly reported in fruit spectroscopy applications [2].

The residual prediction deviation RPD is frequently used in agricultural NIR modeling to assess practical predictive utility [3], [4]. It is calculated as the ratio between the standard deviation (SD) of the reference SSC values in the prediction set and the root mean square error of prediction ($\mathrm{RMSE_p}$) [22]:

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSE_p}} \tag{6}$$

Higher RPD values indicate improved predictive performance. In NIR fruit quality studies, RPD is commonly used as an indicator of model applicability for screening or quantitative analysis.

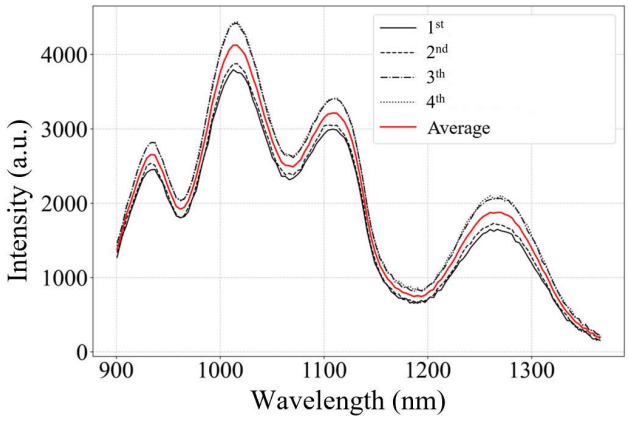
III. Results and Discussion

A. Spectral Consistency Under Consecutive Acquisitions

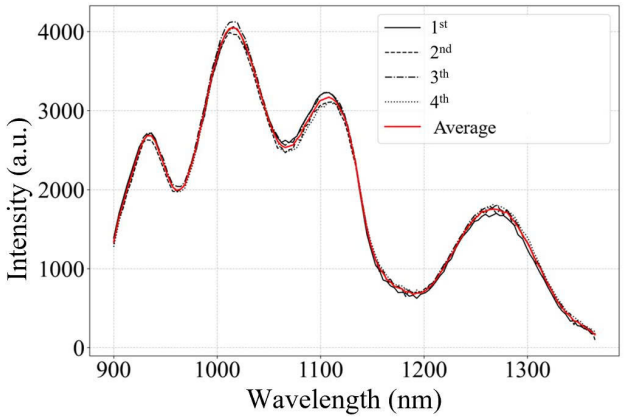
Fig. 5a presents four consecutive interactance spectra acquired at the same nominal measurement region using the manual method. Although the overall spectral profiles follow similar trends, substantial spectral dispersion among consecutive acquisitions is observed across multiple wavelength regions. Pronounced deviations are visible near the major peak around 1000 nm, as well as at secondary local maxima and minima between 1100 and 1300 nm. In several regions, the peak-to-peak separation among scans reaches several hundred a.u., indicating significant variability associated with manual probe repositioning and contact pressure. The calculated RMSD for the manual acquisition mode was approximately 1046 a.u., which represents the absolute spectral dispersion baseline for comparison, reflecting limited spectral consistency under manual operation.

In contrast, Fig. 5b shows four consecutive spectra obtained using the automated acquisition system. The spectra exhibit strong overlap throughout the measured wavelength range, including the local extrema between 1100 and 1300 nm. Only minor deviations are observed among acquisitions. The RMSD for the automated acquisition mode was approximately 363 a.u. Compared with the manual baseline of 1046 a.u., this corresponds to an absolute reduction of 683 a.u. and a relative reduction of approximately 65.3% in spectral dispersion.

These results indicate that the automated system provides improved spectral consistency under consecutive acquisitions, reflecting enhanced acquisition stability. The reduced spectral dispersion supports the reliability of the acquired data and is consistent with the regression performance observed under automated acquisition.



(a)



(b)

Fig. 5. Spectral consistency under consecutive acquisitions at the same nominal measurement region: (a) Manual acquisition and (b) Automated acquisition. Four consecutive scans are shown for each mode; the red curve represents the mean spectrum.

B. Regression Model Performance

Table II presents the calibration and prediction performance of PLS, SVR, and ETR models for mandarin SSC (°Brix) prediction under manual and automated interactance spectral acquisition modes using multiple preprocessing strategies. The objective of this analysis was not to identify an optimal SSC prediction model for grading applications, but rather to evaluate whether the automated interactance acquisition system provides spectral data suitable for chemometric modeling at a level comparable to manual measurement.

1) Effect of regression modeling approaches under different acquisition modes: Clear differences were observed between linear and nonlinear regression methods. Across preprocessing strategies, the PLS model achieved maximum prediction correlations of $R_\mathrm{p} = 0.68$ under manual acquisition and $R_\mathrm{p} = 0.74$

TABLE II. CALIBRATION AND PREDICTION PERFORMANCE OF PLS, SVR, AND ETR MODELS FOR MANDARIN SSC PREDICTION UNDER MANUAL AND AUTOMATED SPECTRAL ACQUISITION MODES USING DIFFERENT PREPROCESSING STRATEGIES.

Model	Acquisition mode	Preprocessing	Calibration			Prediction			
			R_c	R_c^2	$\mathrm{RMSE_c}$	R_p	R_p^2	$\mathrm{RMSE_p}$	RPD
PLS	Manual	MM	0.64	0.41	0.89	0.61	0.37	0.85	1.26
		SNV	0.67	0.45	0.88	0.68	0.38	0.66	1.35
		SG	0.60	0.36	0.92	0.67	0.42	0.84	1.33
		MMADN	0.64	0.41	0.91	0.59	0.27	0.76	1.23
		SG + FD	0.55	0.30	0.94	0.64	0.39	0.94	1.28
		SNV + SG	0.61	0.37	0.95	0.68	0.44	0.66	1.37
		SNV + SG + FD	0.59	0.35	0.94	0.63	0.37	0.79	1.29
		MMADN + SG + MM	0.55	0.30	0.96	0.65	0.36	0.89	1.29
		Unprocessed	0.65	0.42	0.88	0.56	0.26	0.91	1.20
	Automated	MM	0.57	0.33	0.98	0.59	0.32	0.90	1.24
		SNV	0.66	0.44	0.92	0.71	0.39	0.69	1.41
		SG	0.58	0.34	0.96	0.64	0.39	0.91	1.29
		MMADN	0.62	0.38	0.96	0.58	0.24	0.81	1.23
		SG + FD	0.52	0.27	0.99	0.65	0.38	0.96	1.28
		SNV + SG	0.60	0.36	0.98	0.67	0.37	0.71	1.35
		SNV + SG + FD	0.60	0.36	0.96	0.60	0.34	0.87	1.25
		MMADN + SG + MM	0.58	0.34	0.98	0.74	0.42	0.79	1.42
		Unprocessed	0.58	0.33	0.97	0.58	0.32	0.90	1.23
SVR	Manual	MM	0.82	0.67	0.66	0.68	0.46	0.79	1.36
		SNV	0.72	0.49	0.85	0.64	0.33	0.69	1.30
		SG	0.70	0.48	0.83	0.71	0.51	0.78	1.42
		MMADN	0.69	0.47	0.86	0.67	0.38	0.71	1.32
		SG + FD	0.71	0.49	0.81	0.73	0.52	0.83	1.44
		SNV + SG	0.66	0.44	0.89	0.72	0.50	0.62	1.43
		SNV + SG + FD	0.69	0.46	0.86	0.62	0.37	0.79	1.27
		MMADN + SG + MM	0.76	0.57	0.76	0.70	0.48	0.81	1.38
		Unprocessed	0.78	0.59	0.74	0.66	0.43	0.80	1.33
	Automated	MM	0.83	0.67	0.68	0.66	0.43	0.83	1.33
		SNV	0.71	0.49	0.88	0.69	0.44	0.67	1.38
		SG	0.69	0.47	0.86	0.76	0.58	0.76	1.54
		MMADN	0.71	0.48	0.88	0.68	0.43	0.70	1.35
		SG + FD	0.73	0.51	0.82	0.75	0.55	0.82	1.49
		SNV + SG	0.67	0.44	0.92	0.70	0.46	0.66	1.40
		SNV + SG + FD	0.70	0.48	0.87	0.58	0.33	0.87	1.22
		MMADN + SG + MM	0.77	0.58	0.78	0.78	0.56	0.69	1.57
		Unprocessed	0.79	0.60	0.76	0.65	0.41	0.84	1.31
ETR	Manual	MM	0.48	0.23	1.02	0.60	0.30	0.90	1.21
		SNV	0.99	0.96	0.25	0.67	0.27	0.72	1.34
		SG	0.94	0.85	0.45	0.85	0.66	0.65	1.74
		MMADN	0.98	0.93	0.32	0.61	0.27	0.77	1.20
		SG + FD	0.93	0.84	0.45	0.81	0.63	0.73	1.64
		SNV + SG	0.98	0.95	0.27	0.79	0.58	0.57	1.62
		SNV + SG + FD	0.98	0.94	0.29	0.65	0.38	0.78	1.31
		MMADN + SG + MM	0.92	0.82	0.49	0.77	0.55	0.75	1.50
		Unprocessed	0.63	0.38	0.91	0.66	0.34	0.86	1.27
	Automated	MM	0.47	0.22	1.05	0.55	0.25	0.95	1.18
		SNV	0.99	0.97	0.22	0.72	0.33	0.73	1.42
		SG	0.95	0.88	0.41	0.79	0.59	0.74	1.58
		MMADN	0.98	0.95	0.29	0.60	0.28	0.78	1.24
		SG + FD	0.95	0.87	0.42	0.81	0.63	0.74	1.67
		SNV + SG	0.94	0.83	0.50	0.78	0.52	0.62	1.56
		SNV + SG + FD	0.98	0.95	0.27	0.66	0.42	0.82	1.32
		MMADN + SG + MM	0.99	0.96	0.26	0.82	0.59	0.67	1.65
		Unprocessed	0.81	0.61	0.75	0.69	0.39	0.85	1.33

under automated acquisition. In comparison, SVR reached $R_{\mathrm{p}} = 0.73$ (manual) and $R_{\mathrm{p}} = 0.78$ (automated), while ETR achieved $R_{\mathrm{p}} = 0.85$ under manual acquisition and $R_{\mathrm{p}} = 0.82$ under automated acquisition.

These results indicate that nonlinear regression approaches generally provided higher prediction accuracy than the linear PLS model. However, the relative performance between manual and automated acquisition varied depending on the regression method. Automated acquisition yielded higher maximum R_{p} values for both PLS and SVR, whereas ETR exhibited a slight decrease in peak prediction performance compared with the manual mode.

Despite these differences, the overall ranking of regression methods remained consistent across acquisition modes. In both manual and automated measurements, nonlinear models outperformed PLS by a comparable margin. This consistency suggests that the automated acquisition configuration does not alter the fundamental modeling behavior observed with manually acquired spectra.

2) Influence of spectral preprocessing: Preprocessing strategy influenced prediction performance more strongly than acquisition mode. SG was associated with relatively higher R_{p} values in several configurations, particularly for nonlinear regression models. The highest prediction accuracy ($R_{\mathrm{p}} = 0.85$) was obtained using the ETR model with SG preprocessing under manual acquisition. Under automated acquisition, SG-based and combined preprocessing strategies also yielded strong performance, with R_{p} values reaching up to 0.82 for ETR and 0.78 for SVR when combined with additional scaling methods such as MMADN + SG + MM. These observations indicate that spectral smoothing contributes to improved signal stability and can enhance model performance, particularly for nonlinear regression methods.

SNV, commonly used for scatter correction, did not consistently produce higher prediction accuracy compared to smoothing-based approaches. Similarly, scaling procedures such as MM and MMADN affected model performance depending on the regression method and acquisition mode. No single preprocessing strategy uniformly outperformed others across all models, and the effectiveness of preprocessing remained dependent on the interaction between preprocessing method and regression algorithm.

Overall, the results suggest that preprocessing plays a more influential role in determining model performance than acquisition mode, with SG-based approaches providing consistently strong performance across both manual and automated measurements.

3) General synthesis and relevance to study objective: Across all regression approaches and preprocessing configurations, prediction performance under automated acquisition remained within the same range as that obtained under manual acquisition. Maximum R_{p} values and corresponding RPD values were comparable between acquisition modes for each regression method. In several preprocessing–model combinations, particularly PLS and SVR with SG-based preprocessing, slightly higher R_{p} values and lower $\mathrm{RMSE}_{\mathrm{p}}$ values were observed under automated acquisition. Because these differences were small and the objective of this study was to evaluate system applicability rather than superiority, no formal statistical significance analysis was performed. The results therefore support comparable predictive performance between the two acquisition modes.

Most RPD values fell between approximately 1.2 and 1.7, corresponding to exploratory or screening-level prediction capability, as RPD values below 2.0 are generally considered suitable only for rough screening or approximate predictions in NIR fruit quality modeling studies [4], [23]. Although this RPD range indicates limited predictive capability for grading applications [24], achieving grading-level accuracy was not the objective of this study. Overall, the results indicate that automated interactance spectral acquisition provides spectral data of comparable quality to manual measurement, while also offering improved acquisition stability and, in some configurations, slightly enhanced predictive consistency.

The consistent regression behavior and similar prediction ranges across acquisition modes indicate that the automated interactance system provides spectral data comparable to those obtained using manual measurements. Therefore, further improvements in prediction accuracy are more likely to depend on dataset size, sampling diversity, and calibration strategy rather than on acquisition configuration alone.

It should also be noted that this study employed conventional chemometric and machine learning models (PLS, SVR, and ETR) together with commonly used spectral preprocessing techniques to enable a transparent comparison between manual and automated acquisition modes. While these approaches are widely adopted in NIR spectroscopy studies, preprocessing-based modeling pipelines may have limitations in capturing complex nonlinear spectral relationships when larger datasets are available. Recent studies have shown that deep learning approaches, such as convolutional neural networks (CNNs), can automatically learn hierarchical spectral features and reduce reliance on handcrafted preprocessing [2]. With expanded datasets and more diverse sampling conditions, future research may therefore explore deep learning–based spectral modeling to further improve prediction accuracy while leveraging the stable spectral acquisition provided by the proposed automated system.

IV. Conclusion

This study presented and validated an automated interactance NIR spectral acquisition system for non-destructive prediction of mandarin SSC. Compared with the manual setup, the automated system reduced spectral variability among consecutive acquisitions within a measurement session by approximately 65%, confirming improved measurement stability under controlled mechanical positioning and repeatable probe–fruit contact conditions.

Regression models based on PLS, SVR, and ETR were developed using multiple preprocessing strategies under an identical cross-validation protocol to maintain methodological consistency between acquisition modes. Across all model configurations, prediction performance obtained from automated spectra remained within the same performance range as that derived from manual measurements, with several configurations showing slightly higher R_{p} values and lower $\mathrm{RMSE}_{\mathrm{p}}$ values under automated acquisition, particularly for PLS and SVR models combined with SG-based preprocessing.

The relative performance ranking of linear and nonlinear models was preserved across acquisition modes. These results indicate that improved acquisition stability did not adversely affect predictive performance and may be associated with small performance gains in certain configurations, while maintaining overall modeling consistency.

Although most RPD values corresponded to screening-level predictive capability, the primary objective of this study was feasibility validation rather than grading-level optimization. A direct comparison with commercial inline NIR systems in terms of cost, maintenance, and industrial scalability was beyond the scope of the present study; therefore, no quantitative claims are made regarding these aspects. Unlike previously reported online transmission-based systems, the proposed approach emphasizes interactance acquisition with controlled probe positioning and modular mechanical integration. The findings demonstrate that the proposed automated system provides stable spectral data suitable for chemometric modeling and supports future integration into automated fruit grading platforms.

While validated on mandarins, the mechanical positioning and optical acquisition architecture may be adapted to other fruits with similar structural properties. Further improvements in prediction accuracy are therefore more likely to depend on expanded datasets, improved sampling strategies, and advanced calibration methods rather than modifications to the acquisition mechanism alone.

ACKNOWLEDGMENT

This work was funded by the Vietnam Ministry of Education and Training under grant number B2024-TCT-20.

REFERENCES

[1] M. Mei and J. Li, "An overview on optical non-destructive detection of bruises in fruit: Technology, method, application, challenge and trend," *Computers and Electronics in Agriculture*, vol. 213, p. 108195, 2023.

[2] S. Zeng, Z. Zhang, X. Cheng, X. Cai, M. Cao, and W. Guo, "Prediction of soluble solids content using near-infrared spectra and optical properties of intact apple and pulp applying PLSR and CNN," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 304, p. 123402, 2024.

[3] Y. Ozaki, S. Morita, and Y. Morisawa, *Spectral Analysis in the NIR Spectroscopy*. Singapore: Springer Singapore, 2021, pp. 63–82.

[4] B. M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron, and J. Lammertyn, "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review," *Postharvest Biology and Technology*, vol. 46, no. 2, pp. 99–118, nov 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925521407002293

[5] K. B. Walsh, J. Blasco, M. Zude-Sasse, and X. Sun, "Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use," *Postharvest Biology and Technology*, vol. 168, p. 111246, oct 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925521419303230

[6] Y. Yu, Q. Zhang, J. Huang, J. Zhu, and J. Liu, "Nondestructive determination of SSC in Korla fragrant pear using a portable near-infrared spectroscopy system," *Infrared Physics and Technology*, vol. 116, no. January, p. 103785, 2021. [Online]. Available: https://doi.org/10.1016/j.infrared.2021.103785

[7] Z. Hu, Y. Pu, W. Wu, L. Pan, Y. Yang, and J. Zhao, "Online detection of moldy apple core based on diameter and SSC features," *Food Control*, vol. 168, no. August 2024, p. 110879, feb 2025. [Online]. Available: https://doi.org/10.1016/j.foodcont.2024.110879 https://linkinghub.elsevier.com/retrieve/pii/S0956713524005966

[8] X. Tian, S. Fan, J. Li, W. Huang, and L. Chen, "An optimal zone combination model for on-line nondestructive prediction of soluble solids content of apple based on full-transmittance spectroscopy," *Biosystems Engineering*, vol. 197, pp. 64–75, sep 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1537511020301677

[9] Z. Wang, Q. Wang, J. Li, and W. Huang, "Non-destructive detection of soluble solids content in Shawo radish with spatial spectra extraction method based on the full transmission near-infrared spectroscopy," *Journal of Food Composition and Analysis*, vol. 148, no. P2, p. 108364, 2025. [Online]. Available: https://doi.org/10.1016/j.jfca.2025.108364

[10] C. Wang, X. Luo, Z. Guo, A. Wang, R. Zhou, and J. Cai, "Influence of the peel on online detecting soluble solids content of pomelo using Vis-NIR spectroscopy coupled with chemometric analysis," *Food Control*, vol. 167, no. July 2024, 2025.

[11] S. Kim, E. Lee, J. Park, J.-O. Nam, and S. R. Kim, "Evaluating the Nutritional Composition of Unripe Citrus and Its Effect on Inhibiting Adipogenesis and Adipocyte Differentiation," *Journal of Microbiology and Biotechnology*, vol. 34, no. 6, pp. 1206–1213, jun 2024.

[12] K. Timilsina and B. Tripathi, "Chemical quality attributes of mandarin (Citrus reticulata Blanco) as affected by altitude and fruit bearing position in Kavre, Nepal," *Archives of Agriculture and Environmental Science*, vol. 4, no. 3, pp. 319–325, 2019.

[13] R. K. H. Galvão, M. C. U. Araujo, G. E. José, M. J. C. Pontes, E. C. Silva, and T. C. B. Saldanha, "A method for calibration and validation subset partitioning," *Talanta*, vol. 67, no. 4, pp. 736–740, 2005.

[14] X. Wei, J. He, S. Zheng, and D. Ye, "Modeling for SSC and firmness detection of persimmon based on NIR hyperspectral imaging by sample partitioning and variables selection," *Infrared Physics and Technology*, vol. 105, no. October 2019, 2020.

[15] N. F. M. Kasim, P. Mishra, R. E. Schouten, E. J. Woltering, and M. P. Boer, "Assessing firmness in mango comparing broadband and miniature spectrophotometers," *Infrared Physics and Technology*, vol. 115, no. April, p. 103733, jun 2021.

[16] S. Gao and W. Xie, "SSC and pH prediction and maturity classification of grapes based on hyperspectral imaging," *Smart Agricultural Technology*, vol. 8, no. May, p. 100457, aug 2024. [Online]. Available: https://doi.org/10.1016/j.atech.2024.100457 https://linkinghub.elsevier.com/retrieve/pii/S2772375524000625

[17] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, dec 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1568494619302947

[18] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC Trends in Analytical Chemistry*, vol. 132, p. 116045, nov 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165993620302740

[19] M. Fodor, A. Matkovits, E. L. Benes, and Z. Jókai, "The Role of Near-Infrared Spectroscopy in Food Quality Assurance: A Review of the Past Two Decades," *Foods*, vol. 13, no. 21, p. 3501, oct 2024. [Online]. Available: https://www.mdpi.com/2304-8158/13/21/3501

[20] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, apr 2006.

[21] S. Xu, H. Lu, Z. He, and X. Liang, "Non-destructive determination of internal soluble solid content in pomelo using visible/near infrared full-transmission spectroscopy," *Postharvest Biology and Technology*, vol. 214, p. 112990, aug 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925521424002357

[22] P.-L. Nguyen, T.-N. Pham, V.-G.-K. Tran, N.-T. Tran, M. Fukuzawa, C.-N. Nguyen, and C.-N. Nguyen, "Towards Automatic Mango Grading: A Computer Vision System Design for Measuring a Mango's Weight and Skin Defect Areas," *Vietnam Journal of Computer Science*, vol. 12, no. 02, pp. 153–170, may 2025. [Online]. Available: https://www.worldscientific.com/doi/10.1142/S2196888824400025

[23] C. M. Peraza-Alemán, A. López-Maestresalas, C. Jarén, J. I. Ruiz de Galarreta, L. Barandalla, and S. Arazuri, "Mapping acrylamide content in potato chips using near-infrared hyperspectral imaging and chemometrics," *Food Chemistry*, vol. 479, p. 143794, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0308814625010453

[24] L. Kang, H. Luo, X. Ma, J. Yu, H. Liu, H. Liu, and Y. Tong, "Study on the effect of moisture content on the spectral detection of soluble solids in apricot," *Scientific Reports*, feb 2026. [Online]. Available: https://www.nature.com/articles/s41598-026-39890-w