

**The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan**

DEVELOPMENT OF CULTIVAR-OPTIMIZED NIR SPECTROSCOPY MODEL FOR CHERRY TOMATO MATURITY AND SWEETNESS ASSESSMENT

Ssu-Chi Chen¹, Yung-Yi Tsai¹, Joseph Ting², Shu-I Lin³, Shih-Fang Chen^{1,2*}

¹ Department of Biomechatronics Engineering, National Taiwan University, Taiwan.

² Master Program in Global Agriculture Technology and Genomic Science, National Taiwan University, Taiwan.

³ Department of Horticulture and Landscape Architecture, National Taiwan University, Taiwan.

*Corresponding Author: sfchen@ntu.edu.tw

ABSTRACT

"Yunu" cherry tomato cultivars hold substantial commercial value in Taiwan's premium markets, where sweetness serves as a key quality attribute. To enhance cultivar-specific quality assessment, this study evaluates tomato quality in both pre-harvest and post-harvest stages. In the pre-harvest stage, image data were used to establish a Red Ripeness Index (RRI) for evaluating tomato maturity. Color calibration techniques were applied to improve consistency, and the stability and feasibility of the RRI were further examined by analyzing its correlation with hue values under different illumination conditions. In the post-harvest stage, a total of 492 "Yunu" cherry tomato samples were analyzed using near-infrared (NIR) spectroscopy combined with Partial Least Squares Regression (PLSR). The cultivar-optimized model achieved strong predictive performance with an R^2 of 0.8933 and a mean absolute error (MAE) of 0.4697, which was notably more accurate than commercial generic sugar meters (MAE \approx 0.94). The results demonstrate that the proposed RRI provides a reliable tool for ripeness assessment before harvest, while the cultivar-specific NIR model enables accurate and efficient prediction of sweetness after harvest. Together, these outcomes highlight the potential of developing targeted methods to improve both harvest management and post-harvest quality evaluation for premium cherry tomato production.

Keywords: Red Ripeness Index (RRI), Color Calibration, Spectral modeling, Brix prediction, Partial Least Squares Regression (PLSR)

INTRODUCTION

Non-destructive quality assessment is crucial for optimizing harvest timing and post-harvest management in premium tomato production. Traditional approaches often rely on destructive measurements or subjective visual inspection, both of which limit large-scale and consistent application. Effective quality control requires addressing two critical stages: pre-harvest maturity evaluation for harvest scheduling, and post-harvest sweetness assessment for accurate grading and market sorting. Pre-harvest maturity assessment has typically depended on human observation, which is subjective and labor-intensive. Computer vision techniques provide objective alternatives, with color-based indices showing strong potential for ripeness evaluation. In contrast, post-harvest sweetness assessment still relies heavily on destructive sampling, which restricts commercial feasibility. Near-infrared (NIR) spectroscopy enables non-destructive sweetness prediction, and cultivar-specific models can further improve accuracy by accounting for unique biochemical characteristics.

This study targets the “Yunu” cherry tomato, a cultivar valued for its high sweetness in Taiwan’s premium markets, where precise quality control at both harvest and post-harvest stages is essential for maintaining value. We propose a dual-stage framework: (1) color-corrected image analysis through a Red Ripeness Index (RRI) for objective maturity assessment prior to harvest, and (2) NIR spectroscopy with machine learning for post-harvest sweetness prediction and grading.

MATERIALS AND METHODS

SAMPLE ACQUISITION OF NEAR-INFRARED SPECTROSCOPIC DATA

A dataset of 492 “Yunu” cherry tomatoes was collected from Gaoping Tomato and Strawberry Tourist Farm, Hsinchu, Taiwan. Each sample underwent sequential measurements: image acquisition for ripeness evaluation, colorimetric measurements at six positions using a colorimeter (SureColor SD-10; Seiko Epson Corporation), and non-destructive sweetness assessment with a handheld sugar meter (PAL-HIKARi 3 MINi; Atago Co., Ltd.), with measurements taken at four positions per fruit and averaged. The non-destructive measurements were further utilized to evaluate model performance by comparing their predictive consistency against destructive °Brix measurements obtained with a refractometer, thereby quantifying the error between the two approaches.

NIR spectra were collected using a spectrometer (Antaris II; Thermo Fisher Scientific) across 12000-3800 cm^{-1} with 4 cm^{-1} resolution. Each fruit was scanned twice on lateral surfaces with 90° rotation, and spectra were averaged. Final sugar content was measured destructively using a digital refractometer (PAL-1; Atago Co., Ltd.) with three measurements per fruit. The dataset was divided into training (80%) and testing (20%) subsets.

RED RIPENESS INDEX(RRI)

The RRI quantifies tomato ripeness based on HSV color space hue values. The calculation segments the tomato region, computes average hue values excluding black and white pixels (HSV channels at 0 or 180), and applies linear mapping:

$$RRI = 100 - \frac{\sum_{i=1}^N H_i}{N} \times \frac{RRI_{Max} - RRI_{min}}{Hue_{Boundary}} \quad (1)$$

Where H is the Hue value; N is the number of fruit pixels; $Hue_{Boundary}$ is 20; RRI_{Max} and RRI_{min} are 100 and 60. According to this formula definition, the calculated RRI will range between 60 and 100, with higher values indicating that the overall color change of the fruit is closer to red.

MODEL VALIDATION AND STATISTICAL ANALYSIS

RRI reliability was evaluated using 278 controlled lightbox images and 99 greenhouse images across different illumination conditions (Fig. 1a). Since tomatoes exhibit varying ripening levels (approximately 60–100% red maturity) (Fig. 1b), Pearson (r) and Spearman (ρ) correlation coefficients were calculated to assess the relationship between RRI and colorimeter measurements, with comparisons made between original images and those preprocessed using the Perfect Reflection (PR) Algorithm and Multi-Scale Retinex with Color Restoration (MSRCR).

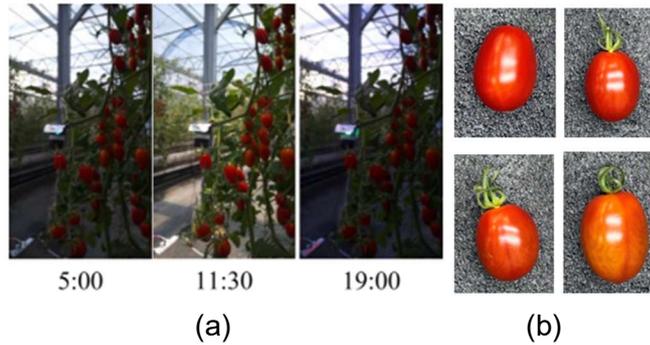


Fig. 1 Cherry tomato images captured under varied conditions.
 (a) Time periods; (b) different ripening levels.

SPECTRAL ANALYSIS

Three preprocessing methods were evaluated: Savitzky-Golay (SG) smoothing, first derivative (SG1), and multiplicative scatter correction (MSC). The optimal preprocessing approach was selected based on cross-validation performance. PLSR modeling linked NIR spectra (X variables) with sugar content (Y variables) through latent variable projection maximizing covariance. Optimal latent variable number was determined by minimizing prediction error while preventing overfitting. Performance metrics includes Mean Absolute Error (MAE), coefficient of determination (R^2), and Residual Predictive Deviation ($RPD = SD/RMSE$) were used for model evaluation.

RESULTS & DISCUSSION

EVALUATION OF RED RIPENESS INDEX PREDICTION

This study examined the association between fruit hue values and the Red Ripeness Index (RRI) to evaluate the reliability of RRI in quantifying ripeness. Three types of image datasets were considered: depth-camera images collected in a controlled lightbox (Type 1), smartphone images collected in a controlled lightbox (Type 2), and smartphone images collected under variable greenhouse conditions (Type 3). Strong negative linear correlations were consistently observed between RRI and measured hue values across all three datasets (Table 1). Type 3 (greenhouse smartphone images) showed the highest correlation ($r = -0.948$ to -0.952), indicating robust performance under variable lighting. Uncorrected images exhibited relatively weak correlations ($r = -0.797$, $\rho = -0.581$) (Fig. 2a), while color-corrected images consistently restored strong negative correlations (Fig. 2b and 2c), demonstrating calibration necessity for stabilizing RRI predictions across variable conditions.

Table 1 The correlation analysis between RRI and Hue values by different data collection methodologies.

Data Type	Coefficient type	RRI Original	RRI MSRCR	RRI PR
Type 1 (175 fruits)	Coefficient (r)	-0.837	-0.827	-0.871
	Coefficient (ρ)	-0.765	-0.749	-0.815
Type 2 (103 fruits)	Coefficient (r)	-0.914	-0.907	-0.913
	Coefficient (ρ)	-0.859	-0.836	-0.854
Type 3 (99 fruits)	Coefficient (r)	-0.948	-0.952	-0.952
	Coefficient (ρ)	-0.931	-0.930	-0.927

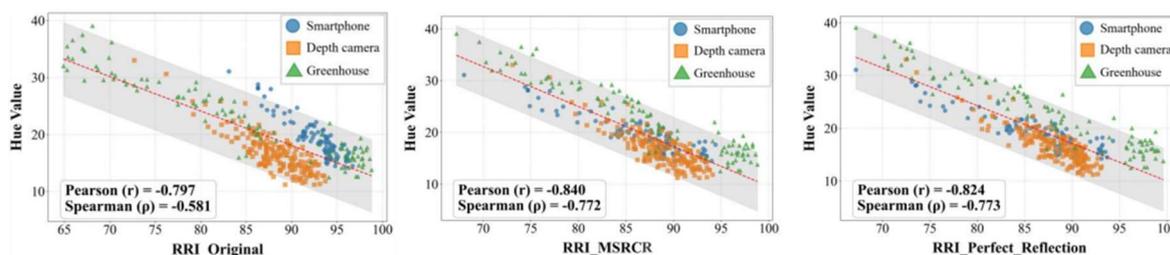


Fig. 2 RRI-Hue relationships of data captured under different environments and devices.

PERFORMANCE OF BRIX PREDICTION

The PLSR model with 25 latent variables achieved excellent predictive performance (Table 2). The testing R^2 of 0.8933 and MAE of 0.4697°Brix significantly outperformed commercial handheld meters (MAE $\approx 0.94^\circ\text{Brix}$). The RPD value of 3.06 exceeds the threshold of 3.0 typically required for practical agricultural applications, confirming model reliability. The superior performance compared to generic sugar meters demonstrates the value of cultivar-specific calibration. The model's accuracy enables reliable quality sorting and harvest optimization for premium "Yunu" tomato production.

Table 2 Performance evaluation of PLSR models for sweetness prediction

Wavelength	Components	Training MAE	Training R^2	Testing MAE	Testing R^2	RPD
3800-12000 cm^{-1}	25	0.41	0.90	0.46	0.89	3.06

CONCLUSIONS

This study successfully developed a dual-stage quality assessment system for "Yunu" cherry tomatoes. The RRI enables reliable pre-harvest maturity evaluation across diverse lighting conditions, while the NIR-PLSR model achieves superior post-harvest sweetness prediction ($R^2 = 0.8933$, MAE = 0.4697°Brix) with practical applicability (RPD = 3.06). These cultivar-optimized methods provide effective solutions for harvest timing optimization and post-harvest grading, supporting premium tomato production quality control.

ACKNOWLEDGEMENTS

This study acknowledges the assistance provided by the staffs at the World Vegetable Center (Tainan City, Taiwan). The research was supported by the National Science and Technology Council, Executive Yuan, Taiwan (Project No. 113-2313- B-002-030-MY3).

REFERENCES

- FAO (Food and Agriculture Organization of the United Nations). 2023. Crops and livestock products. *FAO Corporate Statistical Database*. Available at: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Accessed 24 March 2023.
- Seo, D., B. H. Cho, and K. C. Kim. 2021. Development of monitoring robot system for tomato fruits in hydroponic greenhouses. *Agronomy*, 11(11): 2211.
- Ye, X., T. Doi, O. Arakawa, and S. Zhang. 2021. A novel spatially resolved interactance spectroscopy system to estimate degree of red coloration in red-fleshed apple. *Scientific Reports*, 11(1): 21982.