XU, S., REN, J., LU, H., WANG, X., SUN, X. and LIANG, X. 2022. Nondestructive detection and grading of flesh translucency in pineapples with visible and near-infrared spectroscopy. *Postharvest biology and technology* [online], 192, article 112029. Available from: https://doi.org/10.1016/j.postharvbio.2022.112029

Nondestructive detection and grading of flesh translucency in pineapples with visible and near-infrared spectroscopy.

XU, S., REN, J., LU, H., WANG, X., SUN, X. and LIANG, X.

2022

Nondestructive Detection and Grading of Flesh Translucency in Pineapples with Visible and Near-infrared Spectroscopy

Sai Xu[a,b], Jinchang Ren[d*], Huazhong Lu[a,b*], Xu Wang[c], Xiuxiu Sun[e], Xin Liang[a,b]

a. Institute of Facility Agriculture of Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

b. Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510640, China

c. Institute of Quality Standard and Monitoring Technology for Agro-Products of Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

d. National Subsea Centre, Robert Gordon University, Aberdeen, AB21 0BH, U.K.

e. USDA, Agricultural Research Service, US Pacific Basin Agricultural Research Center, 64 Nowelo Street, Hilo, HI, 96720, USA

* Corresponding author: Huazhong Lu (huazlu@scau.edu.cn), Jinchang Ren (Jinchang.Ren@ieee.org)

Abstract: Rapid, accurate, and nondestructive internal quality detection for large and rough surface fruit, such as translucency in pineapples, is challenging. In this paper, a visible and near infrared (VIS/NIR) spectrum-based platform is proposed for optimized detection of pineapple translucency. The internal quality of three batches of samples harvested at the same maturity but on different dates (early, middle, and mid to late harvest stage) were acquired with different spectral settings: VIS to shortwave NIR (400-1100 nm), NIR (900-1700 nm) and VIS/NIR (400-1700 nm). The pineapple samples were manually cut open and divided into three translucency degrees (no, slight, and heavy), according to marketing standards. The Savitzky Golay (SG) and standard normal variate (SNV) were applied to remove jitter and scattering noise, respectively. The successive projections algorithm, principal component analysis and Euclidean distance were combined for feature extraction and measurement, followed by data modeling using the partial least squares regression and probabilistic neural network (PNN). Data correction, data supplementation, and a combination of these were applied for model updating. Experimental results showed that the optimal solution for pineapple translucency detection was to use 400-1100 nm spectrum with SG, SNV, PNN and data supplementation for model updating. With only the first and second batch of samples used for modeling (validation set accuracy 91.2 %) and updating (validation set accuracy 100 %), the detection accuracy on the third batch samples was 100 %. The proposed methodologies therefore can be used as rapid, nondestructive, and cost-effective tools to detect pineapple translucency to guarantee the marketing of high-quality fruit, which can also guide the postharvest treatment for the pineapple industry to improve market competitiveness as well as to benefit nondestructive quality assessment of other large fruit.

Keywords: Pineapple; translucency; visible and near infrared spectroscopy; nondestructive detection

1. Introduction

Pineapple is one of the most economically important crops in tropical and subtropical areas, however, for the past several decades, it has been damaged by flesh translucency (Paull and Reyes, 1996). Pineapple flesh translucency (PFT) is an irreversible physiological disorder, which affects the flesh and results in low porosity, a water soaked appearance, flat and over-ripe off flavor, rotten taste, and much lower edible quality (Chen and Paull, 2000). Translucent fruit are very susceptible to damage after mechanical collision during transportation and may decay fast during storage (Py et al., 1987). PFT can be caused by either natural or human factors such as pre-harvest temperature, sunburn, excessive rainfall, and overfertilization (Cano Reinoso, 2021; Chen, 1999; Murai et al., 2021; Paull and Chen, 2013). Despite various efforts by researchers and farmers, PFT remains a frequent occurrence at present.

PFT occurs in many countries around the world, such as Benin, Brazil, China, Costa Rica, India, Nigeria, Thailand, USA, etc, and the occurrence rate can reach 87.77% with poor-management (Adetunji et al., 2012; Chen and Paull, 2000; Fassinou Hotegni et al., 2014; Korres et al., 2010; Mandal and Vanlalawmpuia, 2020; Montero-Calderón et al., 2008; Joomwong and Sornsrivichai, 2006). According to the investigation, the PFT occurrence rates in Zhanjiang, the largest producing area of China, were 15 %, 24 %, and 44 % annually from 2019 to 2021. The increasing PFT rate requires the urgent attention of researchers. In China, pineapple is mainly planted by individual farmers without a unified planting standard that makes the control and treatment of PFT even harder in the short term. Thus, it is crucial to explore a nondestructive, fast, and smart solution for PFT detection and grading in order to determine the most suitable postharvest treatments to maintain market quality, protect the brand, and improve market competitiveness.

In the past, acoustic detection was widely applied in industry by knocking the pineapple manually, where a duller sound indicated a more serious degree of translucency. However, the accuracy of acoustic detection is only about 60%, and is also labor intensive and inefficient. Although acoustic impulse-response technique was found feasible for fruit internal quality assessment (Duprat et al., 1997), this technique has not been extensively used by the industry due to its sensitivity to the environmental noise. Haff et al. proposed an X-ray image method for the detection of pineapple translucency, where the detection accuracies for no translucency and extreme translucency were 95 % and 85 %, respectively (Haff et al., 2006). However, X-ray technology has a high cost, is a radiological hazard and is slow, which makes it hard for the agro-product industry to adopt. Therefore, exploration of nondestructive, fast, and cost-effective methods for pineapple translucency detection is still vital and remains unsolved.

Recently, visible and near infrared (VIS/NIR) spectroscopy (Pahlawan et al., 2021), electronic nose (Shi et al., 2018), and machine vision (Naik and Patel, 2017) techniques have become mainstream technologies in nondestructive quality assessment of agro-products. As the occurrence of translucency starts from the heart of the pineapple before spreading out to the whole flesh, electronic nose and machine vision fail to characterize the translucency position of pineapple directly due to the fact that they rely mainly on superficial features such as volatile, color, shape, and size. VIS/NIR spectrum, the previous experiment found, is able to transmit the whole pineapple with enough light intensity and quickly acquire abundant information about the internal quality, which is more suitable for pineapple translucency detection than other technologies (Xu et al., 2021). Additionally, the successful and nondestructive detection of the internal quality of watermelon (Jie and Xuan, 2018) and pomelo (Xu et al., 2020) was demonstrated using the VIS/NIR transmission spectra. However, the nondestructive detection of the internal quality of large fruit is rarely reported. Translucency starts from the middle of the pineapple, which requires an acquired spectrum with a higher signal to noise ratio. Thus, further research on pineapple is still needed due to its unique characteristics.

VIS/NIR spectroscopy has been successfully applied to a wide range of internal quality assessment tasks, such as total soluble solids content, acidity, firmness, pathology, and insect infestation (Adedeji et al., 2020; Li et al., 2016; Lu et al., 2020; Wang et al., 2015), especially for small fruits, such as apple, pear, peach, and kiwi. However, the large size and rough skin of pineapple can easily increase the scattered noise of the spectral signal, leading to difficulty in detecting the internal quality compared to other small fruit. Whether pineapple translucency can be detected using the VIS/NIR spectroscopy or not is still an open question. In addition, most of the existing works focus on nondestructive detection in a lab based setting, where the application cost and model adaptation among different batches of samples are often ignored (Cruz et al., 2021; Zhang et al., 2021). As detection cost and model robustness are two important factors for the potential deployment of the developed techniques in real industrial applications, they need be addressed accordingly.

To tackle these challenging issues, the objectives of the work were to (1) test the efficiency of nondestructive pineapple translucency detection using a developed VIS/NIR spectroscopy-based method; (2) balance the detection accuracy and the cost using two spectroscopies in different wavelengths; (3) explore the robustness of the detection model by using three batches of pineapple samples harvested at different times, one for training, one for updating, and the other for testing in practical applications.

2. Materials and methods

2.1. Pineapple samples

The experimental pineapple samples, variety ‘Bali’, were harvested at Youhao farmland in Zhanjiang, Guangdong Province, China. Three batches of pineapple samples were harvested at different dates for model training, updating, and testing (Table 1). All samples were harvested at the same maturity and~~,~~ with the same cultivation pattern, although they were planted at different dates. Pineapple sampling was conducted in a temporary laboratory near the farmland. For data sampling of the third batch, 10, 20, 20, 20, and 20 samples were used on different postharvest times.

Table 1 Experimental pineapple sample information

Batch number	Sample size	Harvest / sampling date	Data use and data splitting of cross validation
1	100	Both on 16th, April, 2021	(1) Model training: 67 samples (38, 14, and 15 samples were randomly selected from the three categories of no, slight, and heavy translucency,

			respectively) for calibration, the rest 34 samples of batch 1 for validation;
2	100	Both on 1st, May, 2021	(2) Necessity test of model updating: Whole batch 1 for calibration, and whole batch 2 for validation;
3	90	3rd, May, 2021 / 3rd, 5th, 7th, 9th, 11th May, 2021	(3) Model updating: Whole batch 1 plus part (gradually increased) of batch 2 for calibration, the rest of batch 2 for validation; (4) Model testing: Whole batch 1 and part (gradually increased) of batch 2 for calibration, whole batch 3 for validation.

2.2. VIS/NIR based spectrum detection

The lab-based VIS/NIR spectrum detection platform for pineapple internal quality inspection is shown in Figure 1. In consideration of practical requirements of stability of pineapple samples on the assembly line, each pineapple fruit is put on a tray when being imaged. To avoid scattering noise being received by the optical fiber, all light goes through both the input and output optical holes, passing through the pineapple fruit, before being detected by the optical fiber. The whole sampling process was conducted in a dark environment composed of a black box and curtain.

For optimal sampling effect, all the key parameters of the platform were adjustable. The light source is composed of 9 halogen lamps (the power is 100 W each, LM-100, MORITEX Company, Japan), the power of the whole lighting system varies from 100 W to 900 W. The sizes of the input and output optical holes can be determined empirically by changing and testing for multiple times. The distance among the light, pineapple, and optical fiber can also be adjusted by moving components on a sliding rail. In addition, there are only two kinds of commonly used VIS/NIR spectroscopy sensors on the market, namely 400-1100 nm, and 900-1700 nm. Thus, another end of the optical fiber is connected to two different modularized spectroscopy sensors: QE PRO (400-1100 nm) and NIR QUEST (900-1700 nm), both produced by the Ocean Optics Company, USA. The combination of these two can cover 400-1700 nm wavelengths.

After testing repeatedly, the key parameters of the VIS/NIR spectrum detection platform for pineapple inspection were set as follows. The integral times of QE PRO and NIR QUEST were 600 ms and 2000 ms, respectively. The distance between the optical fiber and the tray was 30 mm. The distance between the light and the input optical hole was 84 mm. The power of the light source was 500 W. The pineapple sample was put in the groove in the middle of the tray. The light source, input optical hole, pineapple, output optical hole, and optical fiber were all at the same horizontal level. Under the parameters of the lab-developed VIS/NIR platform, the light would only project onto the pineapple fruit body, before being received by spectroscopy sensors.

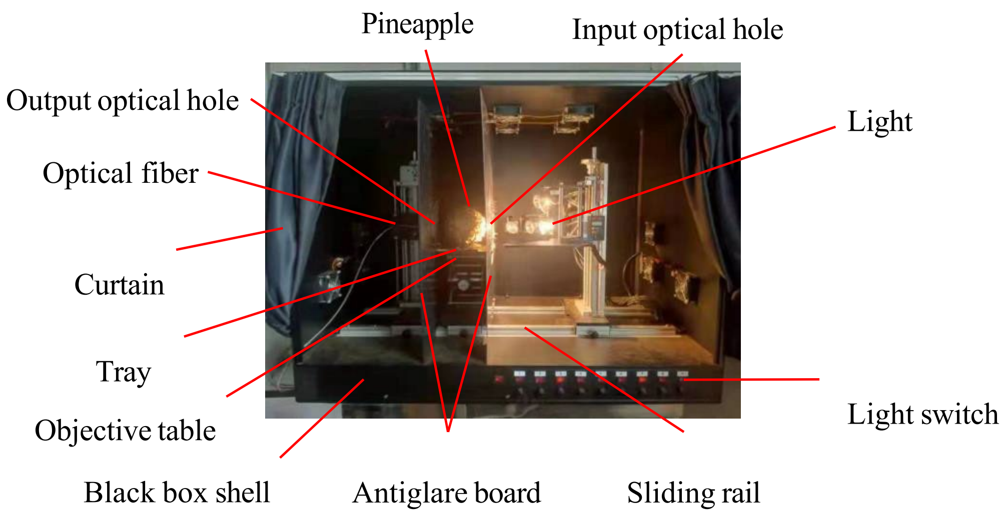

Fig. 1. Structure of lab-developed VIS/NIR spectrum detection platform

2.3. Pineapple translucency degree assessment

After VIS/NIR spectrum sampling, pineapple translucency was assessed based on the spectral data. Pineapples with different degrees of flesh translucency were not obviously different from one another based on visual observation of the skin surfaces. Due to the lack of an industrial standard for pineapple translucency degree assessment, a new pineapple translucency degree assessment method was developed according to practical market considerations. First, the pineapple is cut lengthwise into halves, which are further cut into 12 slices, and then tiled on a table. Second, the translucency degree is evaluated using three categories: no translucency, slight translucency (edibleness and translucent area is up to 10 % of the total sliced area), heavy translucency (edibleness and translucent area is more than 10 % of the total sliced area). Third, the pineapple slices are turned over and checked using Step 2 again, and the most serious degree among the evaluation results was applied to represent the translucency degree of the whole sample.

2.4. Data analysis

2.4.1. Analysis of the first batch of pineapples

The data from the first batch are applied for training the model. The scores from the principal component analysis (PCA) (Wold et al., 1987) were applied for first checking the classification effect and spatial distribution of samples in different pineapple translucency categories. The Savitzky Golay (SG) filter (Press and Teukolsky, 1990) was applied to reduce the jitter noise. The effect of SG is influenced by the order of polynomial and the size of the smoothing window. Standard normal variate (SNV) (Barnes et al., 1989) was applied to reduce the scattered noise due to the rough surface of the pineapple.

After applying SG and SNV as preprocessing, the successive projection algorithm (SPA) (Araújo et al., 2001) was applied for feature extraction, in combination with the PCA and the Euclidean distance (ED) (Danielsson, 1980) as follows. First, all features (transmissivity of each wavelength after SG and SNV processing) were sorted by difference among samples from large to small using SPA. Second, features were gradually added in order (sorted by SPA) from two to the maximum and applied for PCA space classification, respectively; third, ED was applied to calculate the distances between center points of different sample classes in PCA classification space, with the added feature retained if the ED increased (Xu et al., 2015; Xu et al., 2014).

Partial least squares regression (PLSR) (Geladi and Kowalski, 1986) and probabilistic neural network (PNN) (Specht, 1990) were applied for building the detection model. Holdout cross validation was applied for data splitting to avoid overfitting. For model training based on data of batch 1, to avoid inhomogeneity among translucency categories of traditional random data selection of holdout cross validation, 38, 14, and 15 pineapple samples were randomly selected from the no translucency, slight translucency, and heavy translucency categories, respectively, were labeled as 1, 2 and 3, and were selected randomly from the first batch as the calibration set, where the rest of the samples from the same batch were used as the validation set. For PLSR, the factor number (FN) is the variable number selected after feature dimension reduction as the input of modeling, which is the key parameter that affected the detection accuracy, and was determined empirically in this study. In addition, due to the output of the PLSR being decimal, the output was rounded-off to match the labeled value of the translucency degree. For PNN, another key parameter, the Spread value (Ahmadlou and Adeli, 2010), is also empirically determined. The integer output of the PNN can match the labeled translucency degrees.

2.4.2. Analysis of the second batch of pineapples

The data from the second batch of pineapples were used for model updating. To evaluate the adaptability of the detection model to different harvest times, the second batch data (validation set) were applied to test the model which was built based on the first batch data (calibration set). To further improve the adaptability of the detection model, three methods were utilized for comparing the model updating effect, which included data correction, data supplementation, and data correction + supplementation (Candolfi and Massart, 2000; Xie and Ying, 2012). For data correction, a certain number of reference samples were selected from the second batch, while the rest of the samples were used as the validation set. First, the averages of the first batch sample SG and SNV processed spectral data and reference sample SG and SNV processed spectral data of the second batch samples were calculated, to compensate for the difference between averages for each validation samples. For data supplementation, a certain number of reference samples were selected from the second batch, and the rest were used for the validation set. The reference samples were added to the first batch samples to re-train the detection model. For data correction + supplementation, data correction was conducted to compensate for the differences in the averages for all the second batch sample data (reference samples and validation samples), and corrected reference samples were added to the first batch samples to re-train the detection model. To assess the influence of the number of reference samples on the effect of model updating, 5-95 from the second batch ~~o~~f pineapples were randomly selected as the model updating reference samples to add into the first batch of pineapples (calibration set), with a step size of 5, where the rest of the samples (validation set) were used as the validation samples to evaluate the model updating effect.

2.4.3. Analysis of the third batch of pineapples

Data from the third batch of pineapples were applied for model testing (validation set). The refined detection models with three updating methods and a different number of reference samples added into the first batch of pineapples (calibration set) were applied for translucency degree detection on the third batch of data. The results further verify the effectiveness of the updated model.

3. Results and Discussion

3.1. Pineapple translucency degree distribution

The sample distribution in different translucency degrees for the three pineapple batches is shown in Table 2. From the middle of April to the middle of May, 2021, the pineapple translucency occurrence rate first increased and then declined, this same trend can also be found in 2019 and 2020. In China, pineapple translucency only happens in April and May, the season that a large number of pineapples are ready for harvest and marketing~~,~~ due to high rainfall followed by low rainfall in February (Pre-mature stage) in Zhanjiang, Guangdong, China. Previous research also showed that PFT started to occur before the harvest and the occurrence rate increased with maturity development (Chen and Paull, 2000). However, PFT was only slightly related to the harvest season in Thailand, as it occored the whole year (Joomwong and Sornsrivichai, 2006). The reason maybe that Thailand in a tropical area with a high temperature, and intense illumination almost all year.

Table 2 Translucency degree distribution of three batches of pineapples

Batch	No translucency		Slight translucency		Heavy translucency		Sum number
	Number	Proportion (%)	Number	Proportion (%)	Number	Proportion (%)	
1	56	56.0	21	21.0	23	23.0	100
2	24	24.0	31	31.0	45	45.0	100
3	38	42.2	24	26.7	28	31.1	90
Sum	118	40.7	76	26.2	96	33.1	290

3.2. Evaluation of pineapple translucency detection model

3.2.1. Classification using raw spectrum data

Fig. 2(A-C) shows the raw spectrum data from the first batch of pineapples at 400-1100 nm, 900-1700 nm and 400-1700 nm, respectively. As seen, more jitter noise can be observed in the NIR spectrum, especially over 1100 nm. The reason is mainly due to the degradation of degradation of optical energy with increasing wavelength, where the NIR spectrum can be more easily absorbed by water while transmitting the fruit than the VIS wavelengths (Liu et al., 2020). The more translucent fruit has higher transmissivity in the VIS spectrum but lower in the NIR spectrum, because translucency happens with increased water content in the intercellular spaces, thus, VIS transmits better than NIR spectrum.

Pineapple translucency degrees can be classified in three categories based on the PC1 and PC2 of the visible and NIR spectrum within 400-1100 nm (Fig. 2(D)). However, the boundary of each class overlaps with the others, and the clustering performance is poor. For NIR spectrum in 900-1700 nm, the three separate pineapple translucency degrees cannot be differentiated as shown in Fig. 2(E). The occurrence of pineapple translucency is accompanied by changes of flesh color, texture, and other components e.g. sugar accumulation, water content increase, and sugar fermentation (Chen and Paull, 2000). Optically, the visible spectrum and the NIR spectrum are sensitive to color and component (with hydrogen-containing groups) changes (Arendse et al., 2018), respectively, meanwhile both VIS and NIR spectrums are sensitive to flesh texture (Alhamdan et al., 2019). The reason that visible and NIR spectrum within 400-1100 nm have better translucency classification than the NIR spectrum within 900-1700 nm is that translucency involves both flesh color and component changes, to which the visible and NIR spectrum is more sensitive and has a relative high SNR. On the contrary, the NIR spectrum is quite sensitive to component change and a relatively low SNR (Liang et al., 2009). In addition, the classification result using the combined spectrum of 400-1700 nm in Fig. 2(F) is quite similar to those using visible and NIR spectrum only, thus it can be inferred that the primary and most useful information for translucency degree classification is from visible and NIR spectrum within 400-900 nm. In addition, the rough surface brings scattered noise for spectrum detection. Thus, the nondestructive detection of pineapple translucency is too difficult to use methods for fruit internal quality inspection, and requires an improvement in method, from platform to signal preprocessing to modeling.

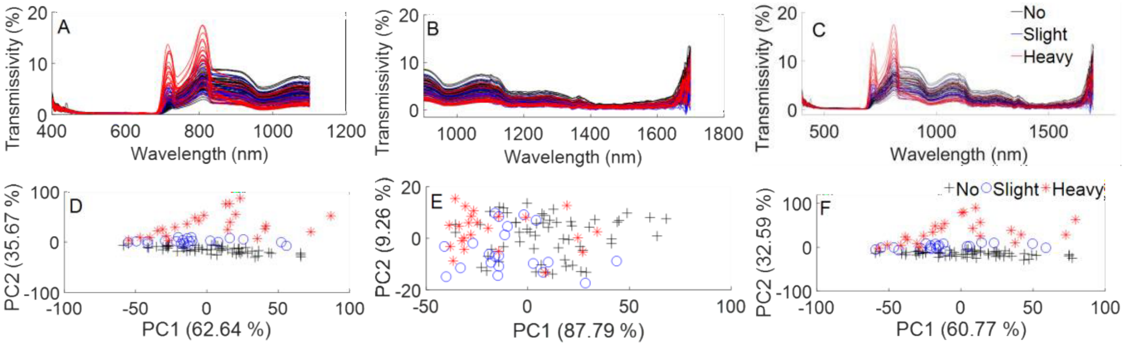

Fig. 2. Raw spectrum data (top) and the PCA space translucency classification results (bottom) of the first batch pineapple based on wavelengths of 400-1100 nm (A, D), 900-1700 nm (B, E), and 400-1700 nm (C, F).

3.2.2. Classification based on denoised spectrum

Considering that the jitter noise in the spectrum may affect the classification, the classification results were evaluated under different SG models for smoothing, where 3 orders 23 points SG, 3 orders 41 points SG, and 3 orders 41 points SG were applied for denoising the spectrums in 400-1100 nm, 900-1700 nm, and 400-1700 nm, respectively for their best performance in the experiments. In addition, due to the highly rough surface of the pineapples, scattering noise is unavoidable in sampling the spectral signal. Thus, SNV was applied to suppress the scattering noise after SG based de-noising, see in Fig. 3(A-C). After applying SG and SNV for denoising, the PC1 and PC2 obtained from different spectrums is visualized, where the results from 400-1100 nm, 900-1700 nm, and 400-1700 nm are shown in Fig. 3(D-F), respectively. The clustering performance is largely improved, and the overlapping is largely reduced when compared to the results from the raw spectral data shown in Fig. 2(D-F). Three translucency degrees are easily distinguishable using 400-1100 nm and 400-1700 nm spectrums, but not using the 900-1700 nm spectrum. However, there is still some overlapping between different translucency classes.

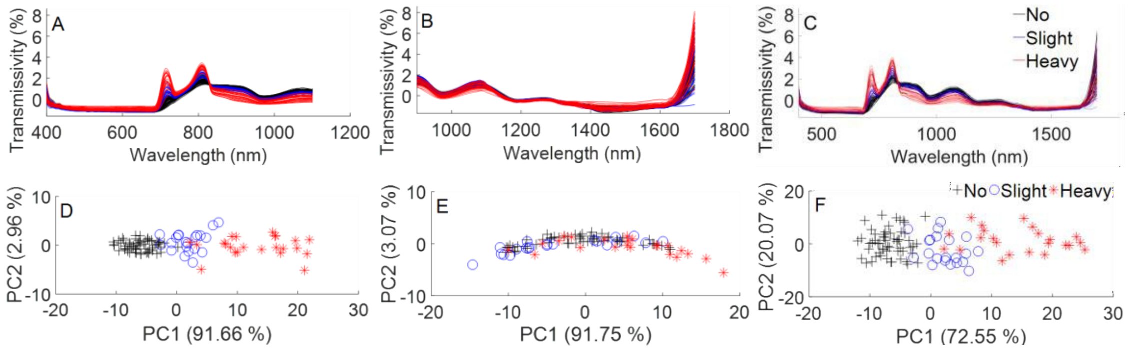

Fig. 3. Raw data (top) and PCA space translucency classification (bottom) of the first batch pineapples after applying SG and SNV for denoising the spectrum of 400-1100 nm (A, D), 900-1700 nm (B, E), and 400-1700 nm (C, F), respectively.

3.2.3. Feature selection and detection model training

Fig. 4 shows the Euclidean distance changes between different translucency classes in the PCA space (composed of PC1 and PC2) while gradually increasing the features in order (ordered by SPA). The effectiveness of this method has been shown in previous research (Xu et al., 2015; Xu et al., 2014). Even the external appearance rarely changed with the occurrence of translucency, however, the color, cell structure, and material composition of internal flesh are apparently different, where all the wavelengths from 400 to 1700 nm can contribute positively to translucency degree detection. In this study, there are 940 and 956 data points in total for 400-1100 nm and 900-1700 nm spectrums, respectively. The test results have shown that these data can be calculated to output the translucency degree within 0.001 sec for detection using SG, SNV and PLSR detection, or 0.06 sec for PNN with SG and SNV on a Lenovo Laptop T14 (Intel i7 CPU, 16.0 GB RAM). Thus, the feature number can satisfy the requirement of real-time detection in industrial applications.

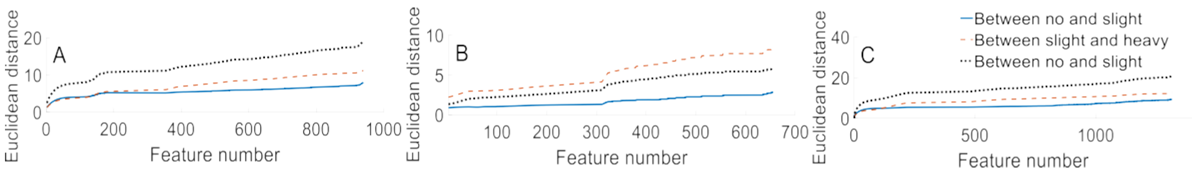

Fig. 4. SPA + PCA + ED contribution analysis of spectral feature from (A) 400 to 1100 nm, (B) 900 to 1700 nm, and (C) 400 to 1700 nm.

Table 3 shows the results of pineapple translucency degree detection using PLSR and PNN, based on different spectra after de-noising with SG and SNV. Pineapple translucency degree can be accurately detected using the 400-1100 nm spectra and 400-1700 nm spectra, but not the 900-1700 nm spectrum. In addition, 400-1100 nm and 400-1700 nm spectra produced the same detection accuracy in the validation set. In general, PNN has better results than PLSR in translucency detection, especially for the validation set. This is because the kernel algorithm of PNN has a stronger capability than the linear regression used in PLSR, especially in modeling the nonlinear characteristics of translucency degree, as seen with the non-linearly separable boundaries between different translucency classes in Fig. 2(D-F). Considering further the relatively low cost of the QE pro for 400-1100 nm compared to the NIR QUEST for 900-1100 nm, the optimal pineapple translucency detection method would use 400-1100 nm spectrum along with SG and SNV for de-noising and PNN for classification.

However, considering the stability in practical use, the testing of the NIR spectrum is still needed for comparison in future experiments.

Table 3 PLSR and PNN based translucency degree detection results using different spectrums with holdout cross validation data splitting

	Wavelength (nm)	Parameter		Calibration set (67 of the first batch) (%)				Validation set (34 of the first batch) (%)			
		FN	Spread	No	Slight	Heavy	Total	No	Slight	Heavy	Total
PLSR	400-1100	11	-	100.0	100.0	93.3	98.5	94.74	100.0	62.5	88.24
	900-1700	11	-	89.5	73.3	66.7	80.6	73.7	57.14	25.0	58.8
	400-1700	14	-	100.0	100.0	100.0	100.0	84.2~~1~~	100.0	93.3	88.24
PNN	400-1100	-	1.2	100.0	100.0	93.3	98.5	100.0	85.7	75.0	91.2
	900-1700	-	0.1	100.0	100.0	100.0	100.0	73.7	28.6	62.5	62.0
	400-1700	-	0.2	100.0	100.0	100.0	100.0	100.0	85.7	75.0	91.2

3.3. Model updating using the second batch of pineapple samples

3.3.1. Necessity test of model updating

Fig. 5(A-B) show the raw spectral data of the second batch of pineapples in 400-1100 nm, and 400-1700 nm, respectively, and their de-noised versions using SG and SNV are shown in Fig. 5(C-D). Compared to the spectral data of the first batch, the second batch has a similar spectral profile, but higher transmissivity in VIS/NIR spectrum and lower transmissivity in NIR spectrum. This is due to the second batch having more serious translucency, as shown in Table 2, which has led to a lower density (or less blockage for VIS/NIR spectrum) and higher water content (more absorption for NIR spectrum) in the flesh.

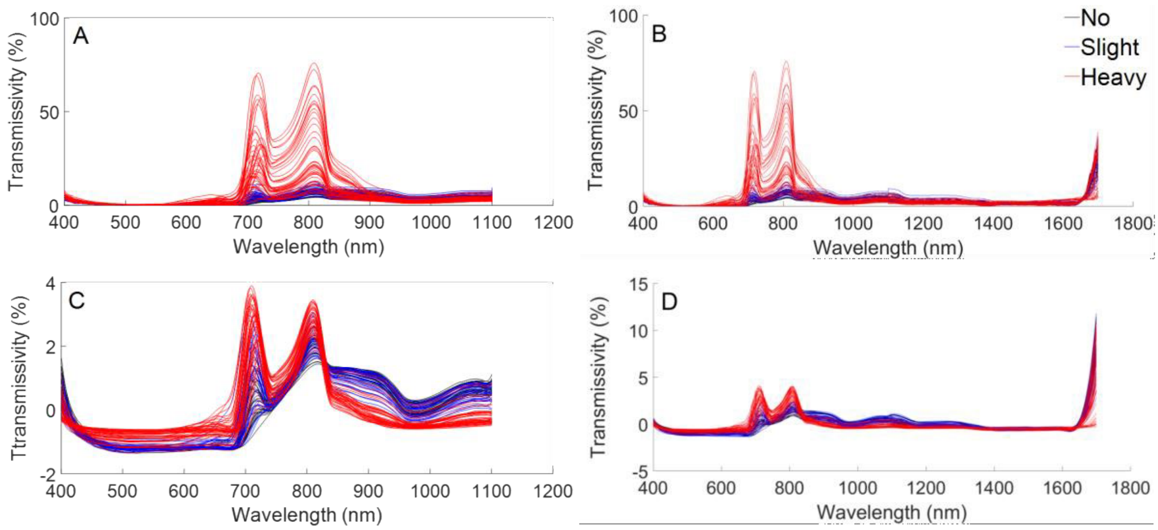

Fig. 5. Raw spectrum data (A, B) and denoised data using SG and SNV (C, D) of the second batch of pineapple samples using 400-1100 nm spectrum (A, C), and 400-1700 nm spectrum (B, D), respectively.

To further validate the efficacy of the trained optimal detection method derived from the first batch of pineapples, the second batch of pineapples were tested using this model. In addition, the samples were also increased for training the model from 67 to 100, all from the first batch of samples, the model was then tested using the 100 samples from the second batch, and the results are summarized in Table 4. As seen, SG + SNV + PNN still produces the best results for the validation set, especially with the VIS/NIR spectrum in the 400-1100 nm spectrum, where the 400-1700 nm spectrum seems unfeasible as it can be easily affected by external disturbance e.g. sample difference. Thus, 400-1700 nm spectra slightly improved the detection accuracy on the calibration set compared to 400-1100 nm spectra (Table 3), which has the risk of lowering the detection accuracy with the addition of interference factors like sample difference (Table 4). Thus, 400-1100 nm spectra already contained major information for pineapple translucency detection, 1100-1700 nm supplied minimal extra useful information compared to 400-1100 nm spectra. In addition, the previous research also proved multi-information fusion had the potential to both increase and reduce detection accuracy (Xu et al., 2019). Thus, 400-1100 nm is a low cost and efficient way for the nondestructive detection of pineapple translucency. Additionally, increasing training samples can improve the detection accuracy on the validation set, but not on the calibration set, as fewer training samples will likely result in over fitting. As the testing was carried out using a different batch of samples, the detection accuracy decreased from 91.2 % to 70 % when compared to testing on the same batch of samples. Thus, model updating is necessary for model adaptation improvement.

Table 4 Results from first pineapple batch detection model used to test the second pineapple batch with holdout cross-validation

	Wavelength (nm)	Modeling sample number	Calibration set (the whole first batch) (%)				Validation set (the whole second batch) (%)			
			No	Slight	Heavy	Total	No	Slight	Heavy	Total
PLSR	400-1100	67	100.0	100.0	93.3	98.5	91.7	19.4	22.2	38.0
		100	100.0	77.8	91.3	92.0	91.7	25.8	37.8	47.0
	400-1700	67	100.0	100.0	100.0	100.0	95.8	6.45	37.8	43.0
		100	96.0	85.2	87.0	91.0	25.0	71.0	62.2	56.0
PNN	400-1100	67	100.0	100.0	93.3	98.5	50.0	45.2	91.1	67.0
		100	100.0	77.8	100.0	94.0	62.5	45.2	91.2	70.0
	400-1700	67	100.0	100.0	100.0	100.0	100.0	3.3	15.6	32.0
		100	100.0	100.0	100.0	100.0	100.0	6.45	17.8	34.0

3.3.2. Translucency detection model updating

The PCA score plot clearly shows the almost original distribution of samples in a same space. To visualize the difference among these three batches of pineapples, the PC1 and PC2 of each SG and SNV processed sample data is shown in Fig. 6 for comparison. Obviously, the first batch cannot cover the other two, which explains the low validation accuracy when testing the second batch of pineapples using the detection model trained on the first batch of pineapples. Model updating research is an essential step to move the application of spectroscopy forward. Thus, model updating is applied to tackle this problem.

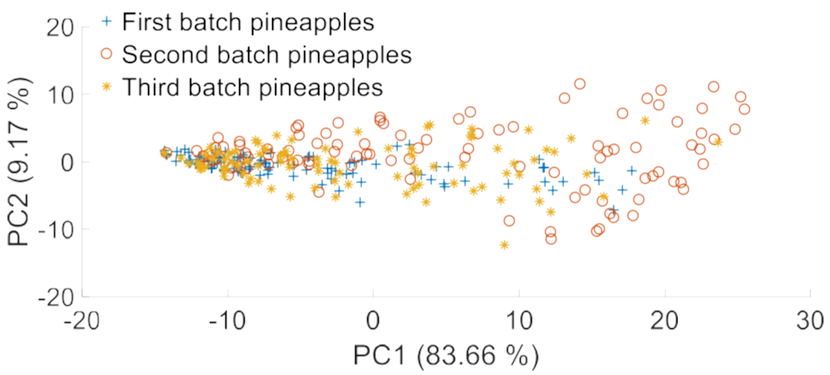

Fig. 6. PCA score plot of the three batches of pineapples

Data correction, data supplementation, and the combination of these two schemes were applied to determine the optimal model updating strategy to tackle the low detection accuracy issue caused by the differences between the samples. Specifically, some of the second batch samples were incrementally taken as references to update the trained model derived from the first batch of pineapple samples. The model updating results are shown in Fig. 7. The detection accuracy on the calibration set of all three model-updating methods increased while adding additional reference samples, which reached over 96 % after five samples (Fig. 7(A)). For the validation set, not surprisingly, the detection accuracy also kept improving (Fig. 7(B)). However, the detection accuracy stability using the data supplementation scheme seems better than the other two methods. With the data supplemented by 40 reference samples from the second batch of pineapples, the detection accuracy on the validation set can be improved from 70 % to 80 %, which is satisfactory for the accurate detection of pineapple translucency degree. For the three model updating methods using 85 reference samples from the second batch of pineapples, the detection accuracy of the validation sets can all achieve 100 %. Both data supplementation (Schimleck et al., 2006) and data correlation (Yao et al., 2010) were proven useful for the different targets' detection model updating. Xie et al. found the model updating effect of data correlation was better than data supplementation in tomato quality detection (Xie and Ying, 2012). Thus, there is no one model updating method which fits all detection targets, due to the specifics of different agricultural products. In addition, a data correlation + supplementation method was proposed in this study, and that model's updating effect was worse than the data supplementation model's effect. However, data correlation + supplementation may be the optimal model updating effect for quality detection of other targets, and future research is required to confirm this. A third batch of pineapples is needed to further validate the efficacy of the three model updating methods.

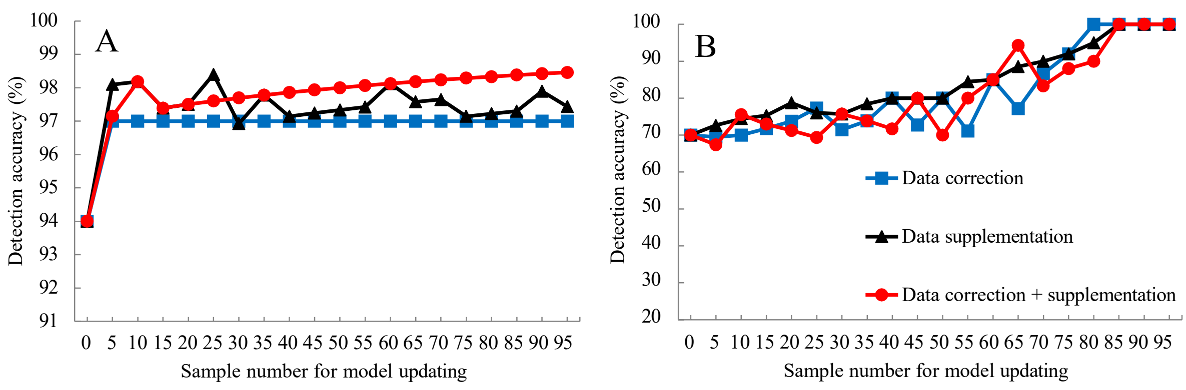

Fig. 7. Results of model updating on the calibration set (A) and the validation set (B) by progressively increasing samples from the second batch.

3.4. Testing on the updated model using the third batch of pineapples

Before testing the accuracy of the updated model using the third batch of pineapples, the average spectra of the first, second, and third batches of pineapples were compared in Fig. 8. The second batch of pineapples shows the highest transmissivity, followed by the third and the first batches. It can also be further confirmed that, pineapples with more serious translucency tend to have higher spectral transmissivity.

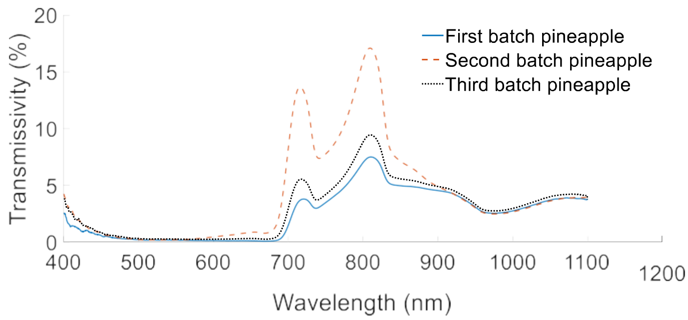

Fig. 8. Average of raw spectrums of the first, second and third batch of pineapples

Ninety pineapple samples in the third batch were harvested to further validate the efficacy of the updated models, using the three model updating schemes for detection of the degree of pineapple translucency. As shown in Fig. 9(A), the data correction based model updating method degrades the validation accuracy with an increasing number of references samples. The reason for this is that each pineapple batch should be more or less different from the others, thus the reference samples of the second batch cannot represent the validation samples from the third batch. In other words, a data correction based model updating method is only suitable for detection within the same batch of training and testing samples. For the data supplementation based model updating method, where the pre-trained model was retrained using the second batch of samples, it can still achieve a very high accuracy for the validation set of the third batch of pineapples for pineapple translucency detection, see in Fig. 9(B). When 40 or more reference samples were added for model updating, the detection accuracies of the total and each individual translucency degree can be at least 80 %, and nearly 100 % when 95 reference samples were added for model updating. Thus the supplementing of the second batch of samples with the first batch can largely improve the data set multiformity to cover the characteristics of the third batch samples. In addition, combining data correction and data supplementation for model updating also shows a poor ability in improving the detection accuracy for practical application, see in Fig. 9(C), and the updating effect on model accuracy on the validation set of the third batch of pineapples seemed unstable with an increasing number of reference samples from the second batch of pineapples for. As a result, data supplementation is found to be the most effective and robust model updating scheme for the detection of pineapple translucency. As shown in Fig. 6, and 8, it can be seen that the first two batches of pineapples can cover most of the characteristics of the third batch. Therefore, the translucency of the third batch can be successfully detected by using the model based on the first two batches of pineapple. Thus, this study provided an optimal VIS/NIR detection platform parameter, and demonstrated the efficiency of SG and SNV for signal preprocessing, and PNN for modeling, for the nondestructive detection of pineapple translucency. Compared to other potential methods, such as acoustic impulse-response technique, the noise interference problem of VIS/NIR spectroscopy, which is more stable and promising for industrial applications, can be efficiently solved by model updating.

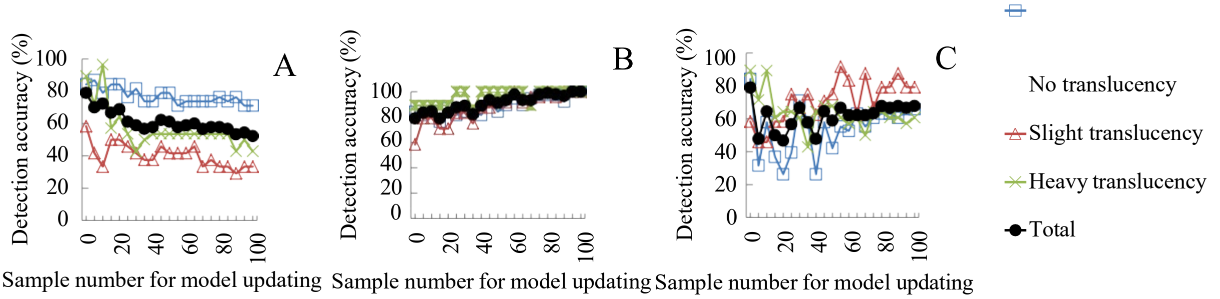

Fig. 9. Detection results on the third batch of pineapples with different model updating approaches including data correction (A), data supplementation (B), and the combined scheme (C).

4. Conclusion

VIS/NIR spectroscopy coupled with data analysis and machine learning was used for nondestructive detection of PFT. The developed low-cost detection platform can meet industrial requirements and work on assembly lines for real-time operations, and the optimal parameters can be empirically determined for the best efficacy.

In this system, SG and SNV were found to be useful in removing the jitter noise caused by low SNR of the large fruit size and the scattering noise caused by the irregular and rough surface of the pineapple. Accordingly, SG and SNV can help to improve the data clustering performance for the classification of pineapple translucency in the PCA space for a reduced dimension of the data. The spectral data from VIS/NIR wavelengths of 400-1100 nm is found particularly useful. Also, PNN produced better results than PLSR in solving nonlinear classification problems of pineapple translucency degrees. To tackle the sample difference of training and validation data, data supplementation is found to produce particularly good results compared to data correction or the combination of these two schemes for model updating. Therefore, the optimal roadmap for pineapple translucency detection in the industry is a recommendation to adopt the 400-1100 nm spectrum data, followed by SG and SNV based de-noising, PNN based data classification, and data supplementation for model updating. The proposed approach can be used as a rapid, nondestructive, and cost-effective framework to detect pineapple translucency. It can help to guarantee the quality of fruit sent to market, to alert pineapple processors around the world to the possible need for post harvest treatment, and to improve the market competitiveness.

Funding

This study has no financial relationship with the organization that sponsored the research.

Declaration of conflicts of interest statement

The authors declare no conflicts of interest.

Acknowledgements

The authors acknowledge the financial support of Laboratory of Lingnan Modern Agriculture Project (NT2021009), Natural Science Foundation of Guangdong Province (2021A1515010834), the Special fund for rural revitalization of Guangdong Province (403-2018-XMZC-0002-90), National Natural Science Foundation of China (31901404), New Developing Subject Construction Program of Guangdong Academy of Agricultural Science (Project No. 202134T), the Presidential Foundation of Guangdong Academy of Agricultural Science (Project No. 202034), Talent Training Program of Guangdong Academy of Agricultural Science (R2020PY-JJX020), Young Talent Support Project of Guangzhou Association for Science and Technology, and Provincial Agricultural Science and Technology Innovation and Extension System Construction Project (2020KJ256). USDA is an equal opportunity provider and employer.

References

Adedeji, A.A., Ekramirad, N., Rady, A., Hamidisepehr, A., Donohue, K.D., Villanueva, R.T., Parrish, C.A., Li, M., 2020. Non-destructive technologies for detecting insect infestation in fruits and vegetables under postharvest conditions: A critical review. Foods 9 (7), 927. https://doi.org/10.3390/foods9070927

Adetunji, C.O., Fawole, O. B., Arowora, K. A., Nwaubani, S.I., Ajayi, E.S., Oloke, J.K., Majolag, 2012. Effects of edible coatings from Aloe vera gel on quality and postharvest physiology of Ananas comosus (L.) fruit during ambient storage. Global Journal of Science Frontier Research Bio-Tech & Genetics 12(5), 39-43. https://journalofscience.org/index.php/GJSFR/article/view/702

Ahmadlou, M., Adeli, H., 2010. Enhanced probabilistic neural network with local decision circles: A robust classifier. Integr. Comput.-Aid. E. 17(3), 197-210. https://doi.org/10.3233/ICA-2010-0345

Alhamdan, A.M., Fickak, A., Atia, A.R., 2019. Evaluation of sensory and texture profile analysis properties of stored Khalal Barhi dates nondestructively using Vis/NIR spectroscopy. J. Food Process Eng. 42 (6), e13215. https://doi.org/10.1111/jfpe.13215

Araújo, M.C.U., Saldanha, T.C.B., Galvao, R.K.H., Yoneyama, T., Chame, H.C., Visani, V., 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometr. Intell. Lab. 57 (2), 65-73. https://doi.org/10.1016/S0169-7439(01)00119-8

Arendse, E., Fawole, O.A., Magwaza, L.S., Nieuwoudt, H., Opara, U.L., 2018. Fourier transform near infrared diffuse reflectance spectroscopy and two spectral acquisition modes for evaluation of external and internal quality of intact pomegranate fruit. Postharvest Biol. Tec. 138, 91-98. https://doi.org/10.1016/j.postharvbio.2018.01.001

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43 (5), 772-777. https://doi.org/10.1366/0003702894202201

Candolfi, A., Massart, D.L., 2000. Model updating for the identification of NIR spectra from a pharmaceutical excipient. Appl. Spectrosc. 54 (1), 48-53. https://doi.org/10.1366/0003702001948105

Cano Reinoso, D.M., 2021. Management Protection on Quality Determination of Flesh Translucency and Fruit Collapse Disease on Indonesian MD2 Pineapple: Universitas Jenderal Soedirman. http://repository.unsoed.ac.id/11056/

Chen, C., 1999. Effects of fruit temperature, calcium, crown and sugar-metabolizing enzymes on the occurrence of pineapple fruit translucencyUniversity of Hawai'i at Manoa. https://www.proquest.com/openview/71e71b7f4d240c048f45ea343d3fe944/1?pq-origsite=gscholar&cbl=18750&diss=y

Chen, C., Paull, R.E., 2000. Sugar metabolism and pineapple flesh translucency. J. Am. Soc. Hortic. Sci. 125 (5), 558-562. https://doi.org/10.21273/JASHS.125.5.558

Cruz, S., Guerra, R., Brazio, A., Cavaco, A.M., Antunes, D., Passos, D., 2021. Nondestructive simultaneous prediction of internal browning disorder and quality attributes in 'Rocha' pear (Pyrus communis L.) using VIS-NIR spectroscopy. Postharvest Biol. Tec. 179, 111562.

https://doi.org/10.1016/j.postharvbio.2021.111562

Danielsson, P., 1980. Euclidean distance mapping. Computer Graphics and image processing 14 (3), 227-248. https://doi.org/10.1016/0146-664X(80)90054-4

Duprat, F., Grotte, M., Piétri, E., Loonisa, D., Studman, C.J., 1997. The acoustic impulse response method for measuring the overall firmness of fruit. Journal of agricultural engineering research. 66 (4), 251-259. https://doi.org/10.1006/jaer.1996.0143

Fassinou Hotegni, V.N., Lommen, W.J.M., van der Vorst, J.G.A.J., Agbossou, E.K., Struik, P.C., 2014. Bottlenecks and opportunities for quality improvement in fresh pineapple supply chains in Benin. International Food and Agribusiness Management Review 17(1030-2016-83025), 139-170. https://doi.org/10.22004/ag.econ.183473

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: A tutorial. Anal. Chim. Acta 185, 1-17. https://doi.org/10.1016/0003-2670(86)80028-9

Haff, R.P., Slaughter, D.C., Sarig, Y., Kader, A., 2006. X - ray assessment of translucency in pineapple. J. Food Process. Pres. 30 (5), 527-533. https://doi.org/10.1111/j.1745-4549.2006.00086.x

Jie, D., Xuan, W., 2018. Review on the recent progress of non-destructive detection technology for internal quality of watermelon. Computers & Electronics in Agriculture 151, 156-164. https://doi.org/10.1016/j.compag.2018.05.031

Joomwong, A., Sornsrivichai, J., 2006. Impact of cropping season in northern Thailand on the quality of smooth cayenne pineapple. II. Influence on physico-chemical attributes. Int. J. Agric. Biol. 8(6), 330-336. https://www.researchgate.net/publication/228728029

Korres, A.M.N., Ventura, J.A., Fernandes, P.M.B., 2010. First report of bacterium and yeasts associated with pineapple fruit collapse in Espírito Santo State, Brazil. Plant Dis. 94(12), 1509-1509. https://doi.org/10.1094/PDIS-04-10-0276

Li, J.L., Sun, D.W., Cheng, J.H., 2016. Recent advances in nondestructive analytical techniques for determining the total soluble solids in fruits: A review. Compr. Rev. Food Sci. F. 15 (5), 897-911. https://doi.org/10.1111/1541-4337.12217

Liang, L., Liu, Z., Yang, M., Zhang, Y., Wang, C., 2009. Discrimination of variety and authenticity for rice based on visual/near infrared reflection spectra. J. Infrared Millim. W. 28 (5), 353-356. https://doi.org/10.3724/SP.J.1010.2009.00353

Liu, Z., Chen, S., Yu, T., Ma, G., Huang, X., Yu, C., Lin, H., 2020. Effects of sunlight on eye. International Eye Science 20 (3), 191-196. https://doi.org/10.18240/ier.2021.01.10

Lu, R., Van Beers, R., Saeys, W., Li, C., Cen, H., 2020. Measurement of optical properties of fruits and vegetables: A review. Postharvest Biol. Tec. 159, 111003. https://doi.org/10.1016/j.postharvbio.2019.111003

Mandal, D., Vanlalawmpuia, C., 2020. Impact of postharvest use of essential oils on quality and shelf life of Indian pineapple. Journal of Postharvest Technology 8(3), 96-105. https://www.researchgate.net/publication/350190874

Montero-Calderón, M., Rojas-Graü, M.A., Martín-Belloso, O., 2008. Effect of packaging conditions on quality and shelf-life of fresh-cut pineapple (Ananas comosus). Postharvest Biol. Tec. 50(2-3), https://doi.org/10.1016/j.postharvbio.2008.03.014

Murai, K., Chen, N.J., Paull, R.E., 2021. Pineapple crown and slip removal on fruit quality and translucency. Sci. Hortic.-Amsterdam 283, 110087. https://doi.org/10.1016/j.scienta.2021.110087

Naik, S., Patel, B., 2017. Machine vision based fruit classification and grading-a review. International Journal of Computer Applications 170 (9), 22-34. https://doi.org/10.5120/ijca2017914937

Pahlawan, M., Wati, R.K., Masithoh, R.E., 2021. Development of a low-cost modular VIS/NIR spectroscopy for predicting soluble solid content of banana, IOP Conference Series: Earth and Environmental Science (Vol. 644, pp. 12047): IOP Publishing. https://doi.org/10.1088/1755-1315/644/1/012047

Paull, R.E., Chen, N.J., 2013. Pineapple translucency and chilling injury in new low-acid hybrids, II Southeast Asia Symposium on Quality Management in Postharvest Systems 1088 (pp. 61-66). https://doi.org/10.17660/ActaHortic.2015.1088.5

Paull, R.E., Reyes, M.E., 1996. Preharvest weather conditions and pineapple fruit translucency. Sci. Hortic.-Amsterdam 66 (1-2), 59-67. https://doi.org/10.1016/0304-4238(96)00905-3

Press, W.H., Flannery, B.P, Teukolsky, S.A., Vetterling, W.T, 1990. Savitzky - Golay smoothing filters. Computers in Physics 4 (6), 669-672. https://doi.org/10.1063/1.4822961

Py, C., Lacoeuilhe, J., Teisson, C., 1987. The pineapple: Cultivation and usesParis: GP Maisonneuve & Larose. http://hdl.handle.net/10524/55419

Schimleck, L.R., Kube, P.D., Raymond, C.A., Michell, A.J., French, J., 2006. Extending near infrared reflectance (NIR) pulp yield calibrations to Newsites and species. J. Wood Chem. Technol. 26 (4), 299-311. https://doi.org/10.1080/02773810601076683

Shi, H., Zhang, M., Adhikari, B., 2018. Advances of electronic nose and its application in fresh foods: A review. Crit. Rev. Food Sci. 58 (16), 2700-2710. https://doi.org/10.1080/10408398.2017.1327419

Specht, D.F., 1990. Probabilistic neural networks. Neural Networks 3 (1), 109-118. https://doi.org/10.1016/0893-6080(90)90049-Q

Wang, H., Peng, J., Xie, C., Bao, Y., He, Y., 2015. Fruit quality evaluation using spectroscopy technology: A review. Sensors-Basel 15 (5), 11889-11927. https://doi.org/10.3390/s150511889

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometr. Intell. Lab. 2 (1-3), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9

Xie, L., Ying, Y., 2012. Calibration model maintenance method for transgenic tomato discriminant. Journal of Jiangsu University (Natural Science Edition) 33(5), 538-542. https://doi.org/10.3969/j.issn.1671-7775.2012.05.009

Xu, S., Lu, H., Ference, C., Qiu, G., Liang, X., 2020. Rapid nondestructive detection of water content and granulation in postharvest “shatian” pomelo using Visible/Near-Infrared spectroscopy. Biosensors 10 (4), 41. https://doi.org/10.3390/bios10040041

Xu, S., Lu, H., Liang, X., Qiu, G., Lin, W., 2021. Effect of artificial ripening on postharvest pineapple fruit quality and its recognition by spectroscopy. Food Science 42 (9), 192-198. https://doi.org/10.7506/spkx1002-6630-20200513-144

Xu, S., Lu, H., Zhou, Z., Lü, E., Jiang, Y., 2015. Identification for guava mechanical damage based on combined hyper-spectrometer and electronic nose. Transactions of the Chinese

Society for Agricultural Machinery 46 (7), 214-219. https://doi.org/10.6041/j.issn.1000-1298.2015.07.031

Xu, S., Sun, X., Lu, H., Zhang, Q., 2019. Detection of type, blended ratio, and mixed ratio of pu′er tea by using electronic nose and Visible/Near infrared spectrometer. Sensors-Basel 19 (10), 2359. https://doi.org/10.3390/s19102359

Xu, S., Zhou, Z., Lu, H., Luo, X., Lan, Y., 2014. Improved algorithms for the classification of rough rice using a bionic electronic nose based on PCA and the Wilks distribution. Sensors-Basel 14 (3), 5486-5501. https://doi.org/10.3390/s140305486

Yao, S., Wu, G., Jiang, Y., Fu, X., Lü, H., Su, M., Pu, J., 2010. Extending hemicelluloses content calibration of acacia spp using nir to new sites. Spectrosc. Spect. Anal. 30 (5), 1206-1209. https://doi.org/10.3964/j.issn.1000-0593(2010)05-1206-04

Zhang, Y., Chen, Y., Wu, Y., Cui, C., 2021. Accurate and nondestructive detection of apple brix and acidity based on visible and near-infrared spectroscopy. Appl. Optics 60 (13), 4021-4028. https://doi.org/10.1364/AO.423994.