Malaysian Journal of Analytical Sciences (MJAS)
Published by Malaysian Analytical Sciences Society

QUALITY ASSESSMENT OF MANGOSTEEN IN DIFFERENT MATURITY STAGES BY HAND-HELD NEAR-INFRARED SPECTROSCOPY

(Penilaian Kualiti Manggis dalam Tahap Kematangan yang Berbeza dengan Genggam Spektroskopi Inframerah Dekat)

Low Shuang Yao[1], Mahmud Iwan Solihin[2], Pavalee Chompoorat[3], Lim Lee Ying[1], Pui Liew Phing[1]*

[1]*Department of Food Science and Nutrition, Faculty of Applied Sciences*
[2]*Dpartment of* Mechanical and Mechatronics Engineering, Faculty of Engineering, Technology and Built Environment,
Faculty of Engineering, Technology and Built Environment
UCSI University, 56000 Kuala Lumpur, Malaysia
[3]*Faculty of Engineering and Agro-Industry,*
Maejo University, Chiang Mai 50290, Thailand.

**Corresponding author: puilp@ucsiuniversity.edu.my*

Received: 11 August 2021 ; Accepted: 18 September 2021; Published: 25 October 2021

Abstract

Quality loss of mangosteen, a tropical fruit, is caused by improper post-harvest handling. A method that can evaluate mangosteen color along with its physical and chemical properties quickly and conveniently should be developed. In this study, 90 mangosteen samples were collected and scanned by the hand-held micro near-infrared (NIR) spectrometer in the wavelength from 900–1700 nm. The mangosteen samples were tested with destructive methods such as color, total soluble solids, reducing sugar content, titratable acidity, and pH to obtain the reference data for the predictive model. Spectral data collected has undergone several pre-processing techniques. The enhanced spectral data are regressed using the regression model. Partial least square (PLS) regression and Principal Component Regression (PCR) are used as the regression method for predicting mangosteen attributes with the help of Orange data mining software. The results showed that color of the sample was not ideal and stage 5 maturity had the lowest L* (27.99), a* (9.03) and b* (7.22). Maturity of stage 6 mangosteen have the highest amount of reducing sugar (10.70×10^{-6} g/100g) and total soluble solids (7.18%). Pearson's correlation was used to determine the relationship between the chemical and physical properties of the mangosteen samples. The PLS and PCR predictive models for reducing sugars were obtained with accuracy as $R^2 = 0.56$ and $R^2 = 0.50$ for the training and testing data, respectively. This achieved accuracy may not be good, but it can be improved in the future by using nonlinear machine learning and ensemble methods such as PLS and PCR. Overall, this study indicated that the NIR spectroscopy technique combined with several pre-processing methods and the predictive model by PLS and PCR could be a rapid and non-destructive method for evaluating the quality and maturity of the mangosteen fruit.

Keywords: hand-held near-infrared spectroscopy, mangosteen, color analysis, total soluble solids, non-destructive quality assessment

Abstrak

Kehilangan kualiti manggis, buah tropika, disebabkan oleh pengendalian pasca tuai yang tidak betul. Kaedah yang dapat menilai warna manggis, sifat manggis fizikal dan kimia dengan kelajuan dan kemudahan yang pantas harus dikembangkan. Dalam kajian ini, 90 sampel manggis dikumpulkan dan diimbas oleh spektrometer mikro dekat inframerah (NIR) genggam. Sampel manggis diuji dengan kaedah yang merosakkan seperti warna, pepejal larut total, pengurangan kandungan gula, keasidan yang dapat ditetrasi dan pH untuk mendapatkan data rujukan untuk model ramalan. Data spektral yang dikumpulkan telah menjalani beberapa teknik pra-pemprosesan. Data spektrum yang ditingkatkan mundur menggunakan model latihan regresi. Regresi Separa Sekurang Persegi (PLS) dan Regresi Komponen Utama (PCR) digunakan sebagai kaedah regresi untuk meramalkan atribut manggis dengan bantuan perisian perlombongan data Orange. Hasil kajian menunjukkan bahawa warna sampel tidak sesuai untuk kematangan tahap 5 yang mempunyai terendah L* (27.99), a* (9.03) dan b* (7.22). Kematangan manggis tahap 6 mempunyai jumlah gula pengurangan tertinggi (10.70 × 10-6 g/100g) dan jumlah perpejal larut (7.18%). Korelasi Pearson digunakan untuk menentukan hubungan antara sifat kimia dan fizikal sampel manggis. Model ramalan PLS dan PCR diperoleh dengan ketepatan masing-masing R^2=0.56 dan R^2=0.50. Secara kesuluruhan, kajian ini menunjukkan bahawa teknik spektroskopi transmisi NIR digabungkan dengan beberapa kaedah pra-pemprosesan dan model ramalan yang dibandingkan dengan PLS dan PCR dapat menjadi kaedah yang cepat dan tidak merosakkan untuk menilai kualiti dan kematangan buah manggis.

Kata kunci: genggam spektroskopi inframerah dekat, manggis, analisis warna, jumlah pepejal larut, penilaian kualiti yang tidak merosakkan

Introduction

Mangosteen (*Garcinia mangostana* L.) is one of the most important tropical fruits in Southeast Asia, and it is under the family of Clusiacaea [1]. The white edible aril of the mangosteen is juicy and has a sweet and slightly sour taste. The peel color of the mangosteen is the major criteria for determining the maturity of the mangosteen. The mangosteen is typically harvested according to color at various stages, from the peel color is greenish yellow with red until blackish purple [2, 3]. Mangosteen is a climacteric fruit. It will begin to ripen as the peel color will turn into a dark purple color after maturation [4]. To obtain the best quality mangosteen, it should be harvested when the peel color is randomly bright with pink to red spots all over the fruit [3]. Near-infrared (NIR) can be used in quantity analysis, where real data was collected to calibrate and validate the predictive model's accuracy in determining the maturity level of mangosteen.

Two types of regression, partial least square (PLS) regression and principal component regression (PCR), were used to construct a model for predicting maturity non-destructive [5]. Near-infrared (NIR) spectrum data from the mangosteen were collected at the maturity level stage 2, stage 5, and stage 6 using a hand-held NIR spectrometer. Several pretreatment methods for the collected NIR spectrum data were applied to eliminate data that were unrelated or irrelevant and incorporated multivariate analysis. These techniques were used to classify the internal attributes, chemical properties, physical properties, and cell structures of the mangosteen [6].

In Malaysia, it is common human visual inspection that is used to determine the maturity stage of mangosteen. The fruit should not be harvested before stage 2 to obtain a high fruit quality [7]. Fruit harvested before stage 2 does not ripen to full flavor, and the excessive latex will flow at the detach point, causing undesirable stains [8]. Destructive techniques require complex sample preparation, which is time consuming [9]. As a result, a fast and accurate method is needed to determine the stage of maturity for mangosteen. NIR was chosen to collect the spectrum data from mangosteen as a near-infrared spectrometer that is characterized by a fast and non-destructive approach. Moreover, nowadays, NIRs are portable because of advances in optical sensor technology.

This study identifies the feasibility to predict mangosteen quality and maturity with a hand-held NIR. This study also determines the physical properties and chemical composition of mangosteen.

Materials and Methods

Raw materials

Mangosteen fruits without any surface defects and with uniform size were obtained from a local farm in Pahang, Malaysia at the end of August 2020. The diameter of the mangosteen fruit ranged from 38 - 62 cm. The mangosteens were categorized into maturity stages by using peel color according to the color index from Tongdee and Suwanagul [8] (Table 1).

There are seven stages defined by the extent of red or purple coloration on the peel. Mangosteens were selected in three different stages, i.e., stage 2, 5, and 6. According to Osman and Milan [10], the harvest standard for mangosteen export in Malaysia was reddish yellow with red spots. Mangosteen at stage 2 in this study displayed a yellowish pink peel with 51–100% scattered pink spots. Meanwhile, mangosteen at stage 5 and 6 were selected to examine the differences in quality assessments. Mangosteen at stage 5 in this study presented with a dark purple color and mangosteen at stage 6 presented with a purple-black color.

Each maturity level was randomly collected for 30 fruits. Mangosteens were cleaned and stored at ambient temperature for analysis. It was then cut in half along the equatorial section after the spectral data was taken. The peels were stored in a ziplock bag. The pulps were ground by the grinder (MX-GM1011, Panasonic, Japan) and filtered by coffee filters to extract the filtrates. The filtrates were stored in the ziplock bag and kept in a freezer for further analysis. Mangosteens at different maturity stages were stored separately for measurement. The samples were kept for 10 days in a storage temperature of 4 - 10 °C.

Table 1. Characteristics of peel color of mangosteen at different maturity stages [8]

Stage	Characteristics of Pericarp
0	Uniformly yellowish-white or with greyish spots
1	Light greenish yellow with scattered pink spots
2	Light greenish yellow or yellowish pink with apparent pink spots all over the fruit
3	Background pinkish, the spots are not as apparent as in stage 2
4	Red or reddish brown with some purple tint
5	Darkened to reddish purple
6	Dark purple to black with little or no red color remaining

Determination of physicochemical properties of mangosteen

Peel color analysis

The color of mangosteens was measured by using Hunter Lab ColorFlex colorimeter (Hunter Associate Laboratory Inc., Reston, USA) [11]. Calibration was done with black and white piles. The value of L*, a* and b* should be 50.87, -25.11 and 14.98, respectively, after calibration.

The sample cup was cleaned with tap water and with no undesirable stain remaining in the cup. A peeler was used to peel the mangosteen's rind, and the peels of mangosteen were placed in the sample cup. The peels of mangosteen completely covered the bottom surface of the sample cup. The sample cup was put into the sensor after it was closed with a cup-like cover to measure the sample and take readings. The button with a lightning bolt would be pressed to reveal the L*, a*, b* values. The L*, a*, b* values were recorded. Chroma (C*) was used to determine the hue's degree of difference compared to grey with the same lightness.

$$C* = \sqrt{a*^2 + b*^2} \quad (1)$$

Hue angle (h*) was used to determine the differences of a certain color in relation to a grey color with the same lightness.

$$h* = \tan^{-1}\left(\frac{b*}{a*}\right) \quad (2)$$

Total soluble solids

The total soluble acid analysis was done using a digital refractometer (MA887, Milwaukee Instruments, Inc., North Carolina, USA) [12]. After the refractometer was calibrated, around two drops of the filtrate sample were put in the refractometer's sample well with a Pasteur pipette. The read button was pressed, and the reading was recorded as total soluble solids (%Brix).

Reducing sugar

Reducing sugar test was done using the Nelson-Somogyi method [13]. Approximately 1 mL of filtrate sample was poured into a test tube and mixed with 9 mL of water. Then, 1 mL of the diluted sample was taken and transferred to another test tube. The diluted sample was added to 9 mL of distilled water for further dilution. Approximately 1 mL of the diluted sample was transferred to the boiling tube and mixed with 1 mL of Nelson mixture.

The mixture would be heated at 100 °C for 20 minutes in the water bath with a boiling tube rack. After heating, the sample was left to cool down on the lab bench. When the sample completely cooled down, the sample was added with 1 mL of arsenomolybdate reagent and 7 mL distilled water. The sample was shaken to make sure the mixture was mixed well. The sample was poured into the glass cuvette. The absorbance of the sample was measured at 510 nm with the spectrometer (UviLine 9400, United Kingdom).

Titratable acidity

The titratable acidity analysis of the pulp was done by titrating sodium hydroxide with the mangosteen sample [12]. Sodium hydroxide with 0.1 M was prepared by dissolving 4 g of sodium hydroxide into 1 L of distilled water. A clamp stand with a pipette was put on the lab bench. The pipette was rinsed with the distilled water and 0.1 M sodium hydroxide. It was then filled with 0.1 M sodium hydroxide until 50 mL and held by the clamp stand. The conical flask was filled with 5 mL of the liquid sample and placed below the pipette. Two drops of phenolphthalein were added to the conical flask as an indicator. The stopcock was twisted open slowly, allowing the 0.1 M NaOH to drop into the conical flash drop by drop until a permanent pink color appeared. The remaining volume of the pipette was recorded for further calculation.

$$\%Titratable\ acidity = \frac{NxVxE}{Wx\ 1000} \text{ x } 100 \quad (3)$$

N was the normality of sodium hydroxide (NaOH), V was the volume made up, E was the acid's equivalent weight, and W was the sample's weight.

pH

The pH of mangosteen was obtained by a pH meter (Metler-Toledo, China) [12]. The mangosteen aril was blended and filtered by a cloth filter. The filtrates were stored in the 30 mL sample bottles. The sample bottle filtrates were measured using a calibrated pH meter with the electrode's bulb fully submerged in the filtrates. The reading of the pH meter was recorded.

NIR spectral data collection

A total of 90 samples were scanned by using the Near Infrared Spectrometer DLP NIRscan Nano EVM (Texas Instrument, USA). The wavelength ranged from 900-1700 nm [15]. The fruit samples were placed directly on the detector window of the instrument. Each sample was scanned at 3 different spots using the reflectance mode. The reflectance mode was the most suitable for mangosteen samples because it has the light level required for radiation to enter the fruit without altering its properties and the reflected light was detected by the detector [16]. The spectra acquired was then averaged and used for further processing [17].

Pre-processing of spectral data

To have a high signal-to-noise ratio, after the spectra of the mangosteen samples were collected, they underwent several pre-processing techniques that eliminated the noise and scattering effects in the spectra data [17]. Cutting, Gaussian smoothing, Savitzky-Golay (SG) smoothing, Multiplicative scatter correction (MSC), and standard normal variate (SNV) were the different techniques used in this study as effective methods of eliminating noise and scattering effects for the spectra data by using the software Orange Data Mining v. 3.26 (Bioinformatics Lab, University of Ljubljana, Slovenia). The spectra (900 - 980.943 and 1621.240–1700) were

cut out, and Gaussian smoothing with a standard deviation of 10 was used to reduce the noise.

Regression models development

All regression training models were done with Orange v. 3.26 (Bioinformatics Lab, University of Ljubljana, Slovenia) [19-22]. Each regression training model had its optimized setting, and each of them was provided with 80% spectra data for training and 20% spectra data for testing. The training model's prediction ability will evaluate the correlation coefficient, and the root means square error between the predicted value and the validation set's measured value. The prediction performance is better when the correlation coefficient is higher, and the root means the square error is lower.

Principal component regression (PCR)

Principal component regression (PCR) is a linear regression model based on principal component analysis (PCA) [23]. PCR regresses the target based on the spectral data matrix, which may reproduce the original data. In PCR, the first step is to decompose data matrix X into a new latent variable by using PCA. The next step, multiple linear regression (MLR), is used between the score obtained in the first step and the interesting property of interest, y. Hence, PCR consists of a linear regression of the scores and y. The linear model of the PCR is expressed as the equation below.

$$Y = Ab + e \quad (4)$$

Partial least square (PLS)

Partial least square (PLS) is a multivariate regression method used in quantitative NIR analysis. PLS is useful when a set of dependent variables from a large set of independent variables must be predicted. PLS is constructing the regression model between two latent variables: the independent X and the dependent Y. The best number of latent variables can be found most commonly through cross-validation based on the minimal prediction error determination.

$$X = TP^T + E \quad (5)$$

$$Y = UQ^T + F \quad (6)$$

T and U are the score matrices for the data set of X and Y, respectively. On the other hand, P and Q are the loading matrices for X and Y, respectively. E and F are the residual matrices [24].

Statistical analysis

All the tests were conducted in triplicate (n=3), and all of the results were presented in the form of mean ± standard deviation. One-way ANOVA was conducted for the comparing means for the three maturity stages of the mangosteen. Tukey HSD test was conducted because the sample size in the experiment is equal. The significant mean differences for all tests were set at 95% interval, which means when $p>0.05$, there were no significant differences. All the data was interpreted by using SPSS Statistics 26 (IBM, Chicago).

Results and Discussion

Color analysis

Color is a significant external quality attribute, and it is often used as a standard for grading the quality of mangosteen [25]. Table 2 shows maturity stage 2 of mangosteen has the highest value of L* (35.39), followed by maturity stage 6 (31.09) and maturity stage 5 (27.99). For a*, there are no significant differences ($p > 0.05$) between stage 2 and stage 6 mangosteen. Maturity stage 5 mangosteen had the lowest value of a* (9.03). For b*, stage 2 mangosteen had the highest value (22.75), and maturity stage 5 had the lowest value (7.22).

L*'s value decreased from maturity stage 2 to maturity stage 5 and increased from maturity stage 5 to maturity stage 6. This means that stage 2 mangosteen had the lightest color and stage 5 mangosteen had the darkest color. The same goes for the a* value and b* value. The a* value and b* values were decreasing from maturity stage 2 to maturity stage 5 and increasing from stage 5 to stage 6. Stage 2 has the reddest and yellowest color among the three maturities. The outer pericarp color of the mangosteen does not meet with the maturity stages of the mangosteen; stage 6 should have the reddest color among the three maturities. The result was contrary to Palapol et al. [26]. The reason may be because the storing of the rind in the freezer for too long. Internal defects of the mangosteen may cause the color on the rind's surface to not match with the proper maturity stage of the mangosteen [27].

Table 2. Color index, %Brix and reducing sugar and acidity of mangosteen in three maturity stages

Stage	L*	a*	b*	Brix (%)	Reducing Sugar Content (g/100g)	Titratable Acidity (%)	pH
2	35.39 ± 4.93 [a]	11.54 ± 3.51 [a]	22.75 ± 3.49 [a]	2.14 ± 0.54 [c]	$5.07 \times 10^{-6} \pm 0.11 \times 10^{-7}$ [c]	2.03 ± 0.41 [a]	3.49 ± 0.18 [c]
5	27.99 ± 2.07 [c]	9.03 ± 1.37 [c]	7.22 ± 2.94 [c]	4.54 ± 0.99 [b]	$8.93 \times 10^{-6} \pm 0.11 \times 10^{-7}$ [b]	1.27 ± 0.26 [b]	3.86 ± 0.18 [b]
6	31.09 ± 2.76 [b]	10.57 ± 1.59 [b]	12.59 ± 2.69 [b]	7.18 ± 0.82 [a]	$10.70 \times 10^{-6} \pm 0.01 \times 10^{-7}$ [a]	0.78 ± 0.21 [c]	4.34 ± 0.45 [a]

*Data collected are calculated in mean ± standard deviations. Different superscripts are indicating the data is significant at $p < 0.05$ by analysis of the Tukey HSD test

Brix and reducing sugar

Brix is usually used to measure the total soluble solids in the fruit, indicating the fruit's sweetness. Table 2 shows the %Brix value of different maturity stages for mangosteen and ranged from 2.14 - 7.18%. The %Brix values in mangosteen with maturity stage 6 were the highest (7.18%); whereas the %Brix values in mangosteen with maturity stage 2 had the lowest (2.14%).

The %brix value increased from maturity stage 2 to maturity stage 6. The results are supported by Castro et al. [28], which has the %brix value ranging from 15.2 - 19.1 due to the prolonged storage and maturation process. %Brix values increased from stage 2 to stage 6, because the accumulated starch in the mangosteen has been converted to sucrose during maturation [29].

Reducing sugar is one of the metrics for determining the maturity level of fruit. The maturity stage 6 mangosteen had the highest amount of reducing sugar content (10.70×10^{-6} g/100 g), and the maturity stage 2 had the lowest amount of reducing sugar content (5.07×10^{-6} g/100 g). The reducing sugar of mangosteen and increase from stage 2 to stage 6 (5.07×10^{-6} g/100 g) and (10.70×10^{-6} g/100 g), respectively. The reducing sugar increased as the maturity level increased, which indicates that the starch accumulated became hydrolyzed into glucose, fructose, and sucrose during the ripening progress [30].

Titratable acidity and pH value

Titratable acidity measured the total concentration of acid in the fruit, and pH values were used to measure hydrogen ionic concentrations in the solution, referred to as the degree of acidity or alkalinity [31]. Both pH and titratable acidity measure the number of organic acids in a fruit [32]. Maturity stage 2 mangosteen had the highest percentage of reducing sugar (2.03), followed by stage 5 mangosteen (1.27) and stage 6 mangosteen (0.78). The mangosteen's pH value of maturity stage 2, 5, and 6 ranged from 3.49 - 4.34. The highest pH of the mangosteen is the maturity stage 6 mangosteen with a value of 4.34. The lowest pH of the mangosteen is the maturity stage 2 mangosteens with a value of 3.49.

The amount of titratable acidity was decreased from 2.03% for stage 2 mangosteens to 0.78% for stage 6 mangosteen. A previous study showed that mangosteen's titratable acidity decreased from 0.74 of maturity stage 1 - 0.60 of maturity stage 4 [3]. The values of pH ranged from 3.49 - 4.34. This showed that an increase of pH corresponded to a decrease in titratable acidity as observed by Islam et al. [33]. The phenomenon was due to the organic acids used in respiration and conversion to sugars during the maturation process. The titratable acidity value was inversely proportional to the pH value. With lower titratable acidity and increases in the pH value at an

elevated maturity level, the mangosteen fruit will be more palatable [34].

Correlation coefficient

Pearson's correlation determined the interdependences of the variable. Table 3 shows the relationship among mangosteen properties such as color (L*, a*, b*), pH, %brix, titratable acidity, and reducing sugars. L* showed a positive correlation to b* and titratable acidity and a* negative correlation to %brix, pH and reducing sugar content. L* represents the intensity of light. L increases indicated that the lightness increases, which means the fruit is immature and immature fruit is acidic and lower in sugar content.

There is a strong positive correlation between pH, %brix and reducing sugar and a strong negative correlation to titratable acidity. This indicates when the titratable acidity increases, pH, %brix and reducing sugar content decreased. The correlation of pH, %brix, reducing sugar content, and titratable acidity has been supported by Ayaz et al. [35]. The explanation for the correlation between pH, %brix, reducing sugar, and titratable acidity were that the organic acids of the fruit will be converted into sugar during ripening, i.e., the sugar content inside the fruit increased [36].

Spectral data

Figure 1 shows the near-infrared reflectance spectra of 90 mangosteen samples. The absorbance of 90 spectral data gathered from the 90 mangosteen samples in the wavelength range from 900–1700 nm. Figure 1 shows the 90 spectral data were the combined wavelength patterns of three maturity stages of mangosteen (Stage 2, 5, and 6).

Figure 1 shows that all the wavelengths can be seen in its distinguished patterns and absorbance number, which varies depending on their maturity stages. The blue wavelengths (maturity stage 2 mangosteen samples) are occupying the top places of the wavelengths as they have the highest absorbance from 900–1700 nm, while the green wavelengths (maturity stage 6 mangosteen samples) occupied the places at the bottom of the wavelengths as it has the lowest absorbance. As for the red wavelength (maturity stage 5 mangosteen samples), they are inconsistent. Most of the red wavelengths have a similar absorbance with the green wavelength, and some have a similar absorbance with the blue wavelengths.

Spectral pre-processing

In this study, pre-processing plays a vital role in improving prediction accuracy and robustness. Spectral data collected will contain noises and unnecessary background information caused by light scattering, over-heated sensors, etc. Noise produced influence and alter the quality information in the spectral data. Hence, it is suggested to pre-process the spectral data before constructing a prediction model.

Five pre-processing techniques will be used in the spectral data. Cutting removes the unnecessary spectral data from the start and the end and only keeps the spectral data's informative part. Savitzky–Golay and gaussian smoothing can reduce or eliminate the noise produced by the background. SNV will remove the scatter effect caused by baseline and pathlength changes with centering and scaling each spectrum. MSC is achieved by regressing each spectrum to an ideal spectrum with the slope and intercept of this linear fit [37]. Figure 2a, b, c, and d, show the changes in the raw spectral data after the four pre-processing methods, start from Cutting followed by Gaussian and Savitzky-Golay smoothing, SNV and EMSC. After several pre-processing uses, the spectral data becomes neat and condense as most of the noises had been removed from the data (Figure 3).

Table 3. Pearson correlation between different attributes, physical and chemical properties

	L*	a*	b*	pH	%Brix	TA
a*	-0.200					
b*	0.709**	0.273**				
pH	-0.219*	-0.120	-0.353**			
%Brix	-0.371**	-0.109	-0.542**	0.723**		
TA	0.382**	0.164	0.562**	-0.664**	-0.792**	
Reducing sugar	-0.480**	-0.241*	-0.692**	0.668**	0.928**	-0.820**

*A double asterisk (**) indicates the correlation is significant at the level of $p \leq 0.01$. A single asterisk (*) indicates the correlation is significant at the level of $p \leq 0.05$. Abbreviations: TA represents Titratable Acidity.

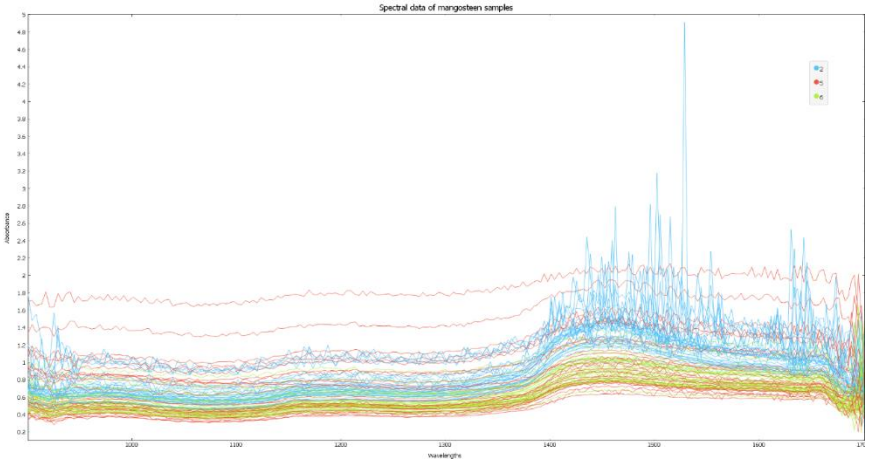
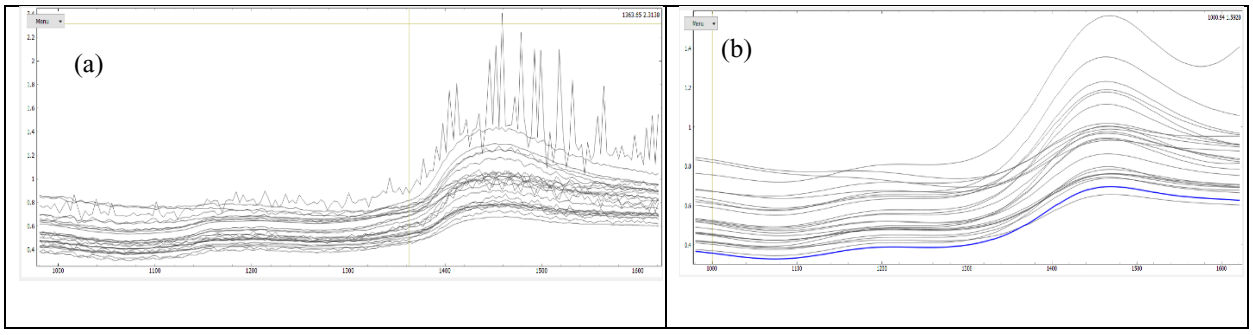


Figure 1. Spectral data collected from the 90 mangosteen samples before pre-processing



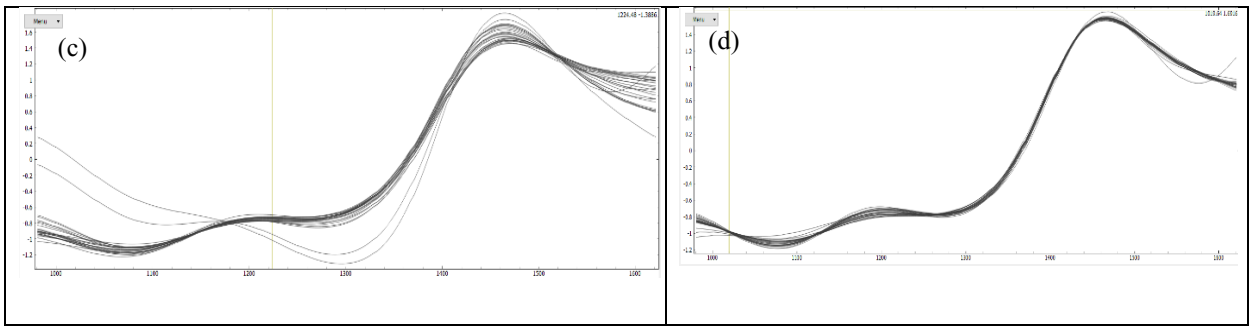


Figure 2. Spectral data after pre-processing; (a) Cut, (b) Smoothing, (c) SNV, (d) EMSC

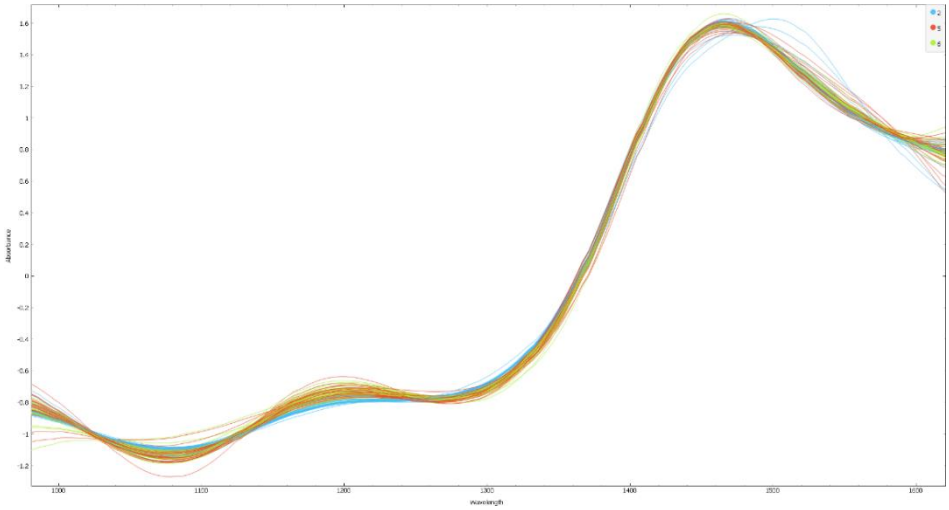


Figure 3. Spectral data after the combination of all pre-processing (Cut, Smoothing, SNV and MSC)

Prediction models for regression

Partial Least Square Regression and Principal Component Regression were used to construct predictive models [38]. The predictive models were developed to acquire information about the attributes of mangosteen buried in the spectral data. Both methods regressed with actual measured value (X-axis) and predicted value from spectra data (Y-axis). Figure 4 shows the graph of predicted and actuals values of the training set of mangosteen's chemical and physical properties with PLS predictive model and PCR predictive model. The pattern of the graph represents the accuracy of the models. The circle mark indicates stage 2, cross indicates stage 5, and triangular indicates stage 6.

From Figure 4a, b, c, d, g, h, i and j, the curve with a slope that is near 45° was compared to the remaining figures. This indicated the predicted value of b, reducing sugar, %brix, and titratable acidity was nearer the actual value [39]. The dot that represented the value for scattering far away from the slopes indicated that the model's predictive performance from PLS and PCR were weak and had a lowered accuracy. The pattern for the scatter graphs was weak, positive, and linear.

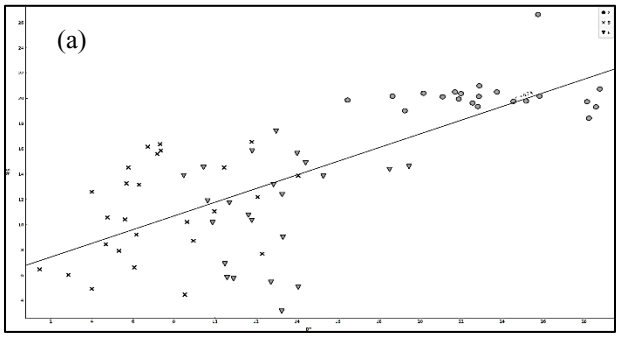
Table 4 shows the p-value = 1 and the standard deviation was high for every property of mangosteen. The predictive model for PLS and PCR was able to determine color, %brix, reducing sugar, titratable acidity, and pH due to $p > 0.05$ in the paired t-test and the high standard deviation showed the data points were away from the mean, which indicated less reliability.

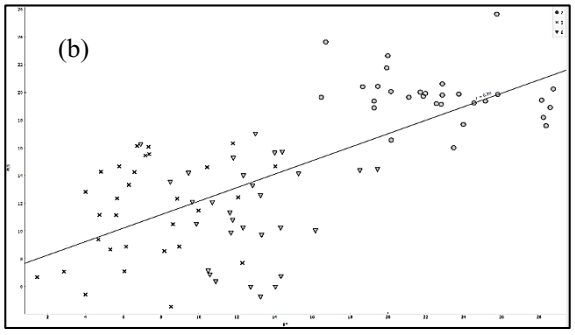
Model evaluation

The indicators for determining the performance prediction models in this experiment were R^2 and RMSE [39]. R^2 is the coefficient of determination, representing the percentages of the explained variance of the result in the dataset. Root mean square error (RMSE) is used to measure the prediction error of the model [41]. The prediction performance will be better when there is a high coefficient of determination, and low root means square error. Table 5 shows the evaluation results for the predictive models using PLS and PCR in determining the attributes of mangosteen.

Table 5 shows L*'s prediction accuracy with the PCR predictive model was very low (R^2 = 0.253). PLS predictive model was slightly better (R^2 = 0.286) than the PCR predictive model but it still was not feasible to predict L*. The prediction results for a* was even lower than L*. The results of PLS and PCR predictive model for a* were (R^2 = 0.155) and (R^2 = 0.165), respectively. The prediction results of the b* with PLS predictive model was over 50% of the coefficient of determination (R^2 > 0.5), but the RMSE was high. It is still not applicable in predicting the b* value as the prediction error was too high. The prediction results of pH for PLS regression were slightly better than PCR regression. The coefficient of determination obtained from PLS for the training set was 0.332 and RMSE of 0.286, while the coefficient of determination obtained from PCR for the testing set was 0.329 and RMSE of 0.390. Hence, the PLS model performed slightly better than PCR in predicting pH.

Moreover, the coefficient of determination for reducing sugar and TSS's training set was 0.563 and 0.325 with RMSE 0 and 1.704 in the PLS predictive model. The PCR predictive model's coefficient's determination was 0.503 in reducing sugar and 0.417 in total soluble solid with RMSE of 0 and 1.648, respectively. PCR predictive model has better predictive performance in determining TSS, and PLS predictive model has better predictive performance in determining reducing sugar. The prediction result of TA in the PLS predictive model was R^2 = 0.440 with RMSE of 0.434, whereas PCR predictive model was R^2 = 0.456 with RMSE of 0.434. PCR predictive model was better at predicting TA. Although PLS had slight advantages in predicting pH and PCR in predicting total soluble solute and TA, they were still not suggested for predicting these attributes due to the low coefficient of determination ($R^2 \leq 0.5$). The only exception is the reducing sugar test as the coefficient of determination was over 50% ($R^2 > 0.5$) and the RMSE was low. Hence, both models are feasible in determining the reducing sugar content in mangosteen; and PLS was a better model for this prediction.





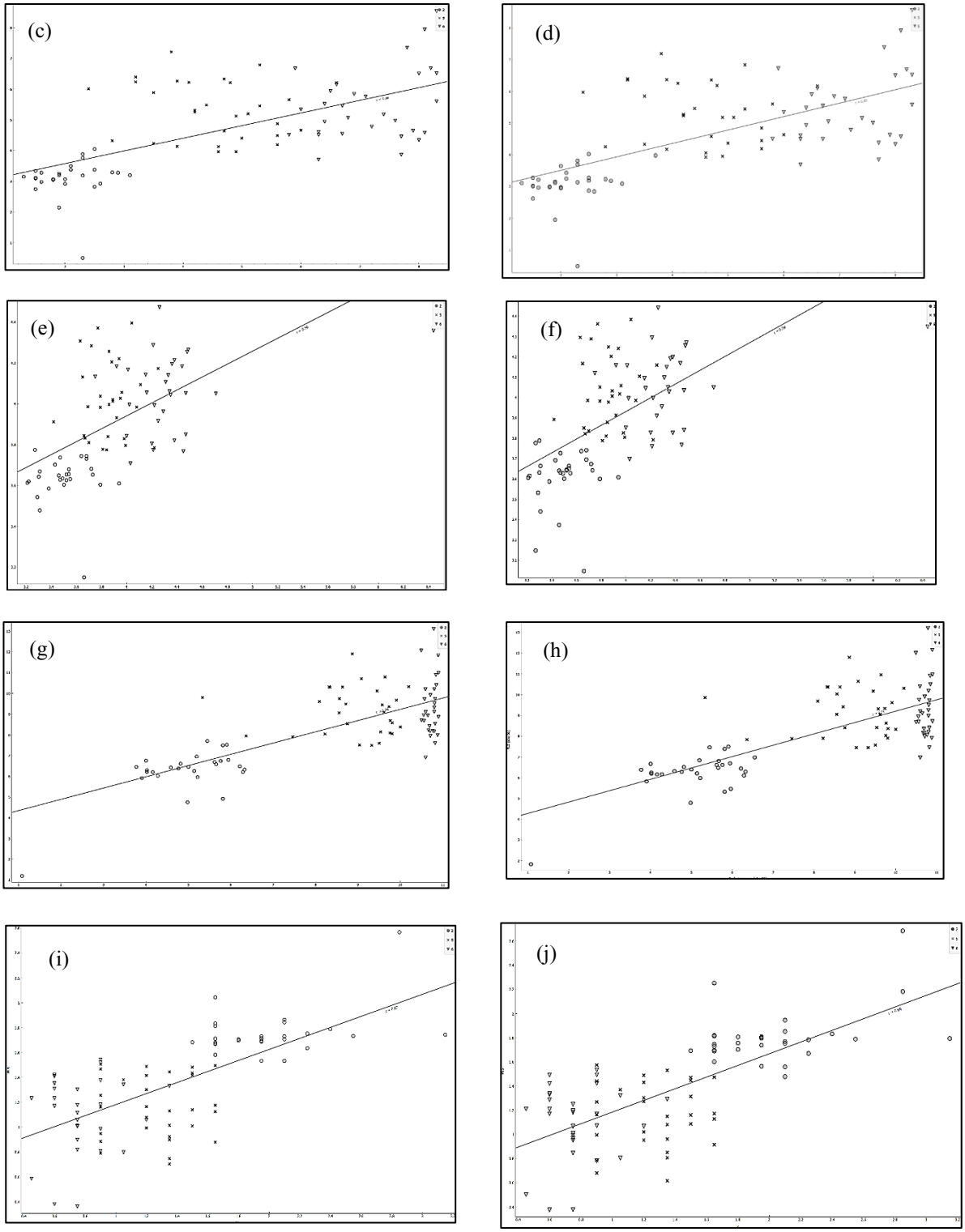


Figure 4. Actual values against predicted values of properties of mangosteen in PLS and PCR predictive model, (a) b* with PLS, (b) b* with PCR, (c) %Brix with PLS, (d) %Brix with PCR, (e) pH with PLS, (f) pH with PCR, (g) reducing sugar with PLS, (h) reducing sugar with PCR, (i) titratable acidity with PLS, and (j) titratable acidity with PCR

Table 4. Actual and predicted values of color, %brix, reducing sugar content, titratable acidity and pH

Value	L*	a*	b*	Brix (%)	Reducing Sugar Content (g/100 g)	Titratable Acidity	pH
Actual	31.49 ± 4.59	10.38 ± 2.56	14.18 ± 7.15	4.62 ± 2.22	$8.24 \times 10^{-6} \pm 2.5 \times 10^{-6}$	1.36 ± 0.60	3.89 ± 0.05
Predicted	31.49 ± 2.31	10.38 ± 1.04	14.18 ± 4.98	4.62 ± 1.44	$8.24 \times 10^{-6} \pm 1.88 \times 10^{-6}$	1.36 ± 0.35	3.89 ± 0.03
P-value	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Data collected are calculated in mean ± standard deviations. P-values are determined by analysis of Paired T-test

Table 5. Regression results for attributes of mangosteen by using PLS and PCR

Quality Attribute	Methods	R^2 (Training & Testing)	RMSE (Training & Testing)
L*	PLS regression	0.286 & 0.245	3.857 & 4.007
	PCR regression	0.253 & 0.291	3.950 & 3.738
a*	PLS regression	0.155 & 0.143	2.315 & 2.347
	PCR regression	0.165 & 0.127	2.323 & 2.379
b*	PLS regression	0.542 & 0.500	4.931 & 5.020
	PCR regression	0.486 & 0.484	5.095 & 5.046
pH	PLS regression	0.332 & 0.436	0.286 & 0.545
	PCR regression	0.329 & 0.353	0.390 & 0.267
Reducing Sugar	PLS regression	0.563 & 0.495	0.000 & 0.000
	PCR regression	0.503 & 0.727	0.000 & 0.000
Total soluble solids (TSS)	PLS regression	0.409 & 0.503	1.704 & 1.530
	PCR regression	0.417 & 0.478	1.648 & 1.172
Titratable Acidity (TA)	PLS regression	0.440 & 0.599	0.434 & 0.340
	PCR regression	0.456 & 0.669	0.434 & 0.345

*Abbreviations: PLS represents Partial Least square; PCR represents Principal Component Regression; R^2 represents correlation coefficient; RMSE represents Root Mean Square Error.

Conclusion

Nowadays, methods in determining the quality of mangosteen fruit are usually destructive. Farmer and the mangosteen industry workers often harvest the mangosteen at the wrong maturity, which produces poor quality fruit that can't be exported. Hence, NIRS was used as one of the significant non-destructive techniques to determine quality and the maturity stages of the fruit. In this study, the physical and chemical properties of mangosteen were tested. NIRS coupled with pre-processing, PLS and PCR predictive modal were able to feasibly identify the maturity stage of mangosteen by determining the reducing sugar content. From the present results, mangosteen attributes such as %brix, reducing sugar, pH, and titratable acidity were followed by a ripening process for the mangosteen. Brix, reducing sugar, and pH increased while titratable acidity decreased along with the ripening process as sugar content inside the mangosteen increased due to the sugar conversion of organic acid. The maturity stage 5 mangosteen had the darkest color, whereas maturity 2 mangosteen had the lightest color. The %brix of the mangosteen samples ranged from 2.14–7.18%. Maturity stage 2 mangosteen revealed the lowest reducing sugar content ($5.07 \times 10^{-6} \pm 0.11 \times 10^{-7}$) and pH (3.49). In addition, maturity stage 2 mangosteen showed the highest titratable acidity (2.03%) among the three different maturity mangosteens. However, maturity

stage 6 mangosteen had the lowest titratable acidity (0.78%). In conclusion, NIRS was only feasible in determining the reducing sugar content with acceptable accuracy (R^2 > 50%) in both predictive models. This achieved accuracy may not be excellent, but it can be improved in the future by using nonlinear machine learning and ensemble methods as PLS and PCR.

Acknowledgements

The authors would like to express the gratitude to the Centre of Excellence for Research, Value Innovation, and Entrepreneurship (CERVIE) – UCSI University for supporting this project, under PSIF project code: Proj-2019-ln-FETBE-063.

References

1. Ayman, E. L., Hassan, S. M. and Osman, H. E. H. (2019). Mangosteen (*Garcinia mangostana* L.). In Nonvitamin and Nonmineral Nutritional Supplements. Academic Press, Massachusetts: pp. 313-319.
2. Ketsa, S. and Paull, R. E. (2011). Postharvest biology and technology of tropical and subtropical fruits, ed. E.M. Yahia, Woodhead Publishing Series in Food Science, Technology and Nutrition, Amsterdam: pp. 1-32.
3. Sumiasih, I. H., Poerwanto, R. and Efendi, D. (2019). Study of several stages of maturity and storage temperature on color changes and shelf life of mangosteen (*Garcinia mangostana* L.). *International Journal of Applied Biology*, 3(1): 45-54.
4. Manurakchinakorn, S., Chainarong, Y. and Sawatpadungkit, C. (2016). Quality of mangosteen juice colored with mangosteen pericarp. *International Food Reseach Journal*, 23: 1033-1034.
5. Jha, S. N., Narsaiah, K., Jaiswal, P., Bhardwaj, R., Gupta, M., Kumar, R. and Sharma, R.. (2014). Non-destructive prediction of maturity of mango using near infrared spectroscopy. *Journal of Food Engineering*, 124: 152-157.
6. Hayati, R., Munawar, A. A. and Fachruddin F. (2020). Enhanced near infrared spectral data to improve prediction accuracy in determining quality parameters of intact mango. *Data in Brief*, 2020: 105571-105572 .
7. Yahia E. M. (2011). Post-harvest biology and technology of tropical and subtropical fruits: mangosteen to white sapote. Elsevier, Amsterdam.
8. Tongdee, S. C. and Suwanagul A. (1989). Post-harvest mechanical damage in mangosteens. *ASEAN Food Journal Malaysia*.
9. Prabasari, I., Utama, N. A., Hasanah, N. A. U., Riyadi, S. and Hariadi, T. K.. (2016). Non-destructive method for maturity index determination of *Garcinia mangostana* L. using image processing technology. *International Society for Horticultural Science*, pp. 1-10.
10. Osman, M. B. and Milan, A. R. (2006) Mangosteen: *Garcinia mangostana* L. Southampton Centre for Underutilised Crops. University of Southampton, Southampton, UK.
11. Chang, L. S, Tan, Y. L., and Pui, L. P. (2020). Production of spray-dried enzyme-liquefied papaya (*Carica papaya* L.) powder. *Brazilian Journal of Food Technology*, 23: e2019181.
12. Pui, L. P., Karim, R., Yusof, Y. A., Wong, C. W. and Ghazali, H. M. (2018). Physicochemical and sensory properties of selected 'cempedak' (*Artocarpus integer* L.) fruit varieties. *International Food Research Journal*, 25(2): 861-869.
13. Sadasivam, S. (1996). Biochemical Methods. New age international, India.
14. AOAC. (2000). Official Methods of Analysis of AOAC International, 17th edn. Association of Official Analytical Chemists, Arlington.
15. Abdullah Al-Sanabani, D. G., Solihin,M. I., Pui, L. P., Astuti, W., Ang, C. K. and Hong, L. W. (2019). Development of non-destructive mango assessment using handheld spectroscopy and machine learning regression. *Journal of Physics: Conference Series,* 1367(1): 012030.
16. Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., Dupont, P., Rondia, A. and Lateur, M. (2013). Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, 93(2): 238- 244.

17. Yan, H., Xu, Y. C., Siesler, H. W., Han, B. X. and Zhang, G. Z. (2019). Hand-held near-infrared spectroscopy for authentication of fengdous and quantitative analysis of mulberry fruits. *Frontiers in Plant Science*, 10: 1548- 1549.
18. Wang, H., Peng, J., Xie, C., Bao, Y. and He, Y. (2015). Fruit quality evaluation using spectroscopy technology: a review. *Sensors*, 15: 11889-11927.
19. Solihin, M. I., Zekui, Z., Ang, C. K., Heltha, F. and Rizon, M. (2021). Machine learning calibration for near infrared spectroscopy data: A visual programming approach. Springer, Singapore: pp. 577-590.
20. Irfan, U. Bin, Pui, L. P. and Solihin, M. I. (2020). Feasibility study of detecting palm oil adulteration with recycled cooking oil using a handheld NIR spectroscopy. *AIP Conference Proceedings*, 2306(1): 020019.
21. Ting, D. F., Pui, L. P. and Solihin, M. I. (2020). Feasibility of fraud detection in milk powder using a handheld near-infrared spectroscopy. *AIP Conference Proceedings*, 2306(1): 020017.
22. Liew, K. T., Pui, L. P. and Solihin, M. I.. (2020). Feasibility of fraud detection in rice using a handhelf near-infrared spectroscopy. *AIP Conference Proceedings*, 2306(1): 020018.
23. Hanrahan, G.. (2009). Modelling of pollutants in complex environmental systems. ILM Publications, USA.
24. Salleh, N. A. M. and Hassan, M. S. (2019). principal component regression (PCR) and linear support vector machine regression (SVMR). *Journal of Physics: Conference Series*, 1366: 1- 12114.
25. Yahaya, O. K. M., MatJafri, M. Z., Aziz, A. A. and Omar, A. F. (2014). Non-destructive quality evaluation of fruit by color based on RGB LEDs system. *In 2014 2nd International Conference on Electronic Design (ICED)*: pp. 230-233.
26. Palapol, Y., Ketsa, S., Stevenson, D., Cooney, J. M., Allan, A. C. and Ferguson, I. B. (2009). Color development and quality of mangosteen (*Garcinia mangostana* L.) fruit during ripening and after harvest. *Postharvest Biology and Technology*, 51(3): 349-353.
27. Jarimopas, B., Pushpariksha, P. and Singh, S. P. (2009). Postharvest damage of mangosteen and quality grading using mechanical and optical properties as indicators. *International Journal of Food Properties*, 12(2): 414-426.
28. Castro, M. F. P. P. M., Anjos, V. D., Rezende, A. C. B., Benato, E. A. and Valentini, S. R. T. (2012). Post-harvest technologies for mangosteen (*Garcinia mangostana* L.) conservation. *Journal of Food Science and Technology*, 32: 668-672.
29. Yamaki, S. (2010). Metabolism and accumulation of sugar translocated to fruit and their regulation. *Journal of the Japanese Society for Horticultural Science*., 79: 1-15.
30. Tadesse, T., Workneh, T. S. and Woldetsadik, K. (2012). Effect of varieties on changes in sugar content and marketability of tomato stored under ambient conditions. *African Journal of Agricultural Research*, **7**(14): 2024-2030.
31. Trick, J. K., Stuart, M. and Reeder, S. (2008). Contaminated groundwater sampling and quality control of water analyses. *Environmental Geochemistry*, 29-57.
32. Susan, F. (2016). Complete Course in Canning and Related Processes. Woodhead, USA.
33. Islam, M., Khan, M. Z. H., Sarkar, M. A. R., Absar, N. and Sarkar, S. K. (2013). Changes in acidity, TSS, and sugar content at different storage periods of the post-harvest mango (*Mangifera indica* L.) influenced by Bavistin DF. *International Journal of Food Science*, 2013: 1-8.
34. Rooban, R., Shanmugam, M., Venkatesan, T. and Tamilmani, C. (2016). Physiochemical changes during different stages of fruit ripening of climacteric fruit of mango (*Mangifera indica* L.) and non-climacteric of fruit cashew apple (*Anacardium occidentale* L.). *Journal of Applied and Advanced Research*, 1(2): 53-58.
35. Ayaz, F. A., Kadioglu, A. S. İ. M., Bertoft, E., Acar, C. and Turna, I. (2001). Effect of fruit maturation on sugar and organic acid composition in two blueberries (*Vaccinium arctostaphylos* and *V. myrtillus*) native to Turkey. *New Zealand Journal of Crop and Horticultural Science*, 29: 137-141.

36. Batista-Silva, W., Nascimento, V. L., Medeiros, D. B., Nunes-Nesi, A., Ribeiro, D. M., Zsögön, A. and Araújo, W. L.(2018). Modifications in organic acid profiles during fruit development and ripening: correlation or causation?. *Frontiers in Plant Science*, 9: 1689-1690.
37. Huang, J., Romero-Torres, S. and Moshgbar, M.. (2010). Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review*, 2010: 2-10.
38. Wentzell, P. D. and Montoto, L. V. (2003). Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and Intelligent Laboratory Systems*, 65(2): 257-279.
39. Jaiswal, P., Jha, S. N., and Bharadwaj, R. (2012). Non-destructive prediction of quality of intact banana using spectroscopy. *Scientia Horticulturae*, 135: 14-22.
40. Richter, K., Hank, T. B., Atzberger, C. and Mauser, W. (2011). Goodness-of-fit measures: what do they tell about vegetation variable retrieval performance from Earth observation data. *Remote sensing for agriculture, ecosystems, and hydrology XIII. International Society for Optics and Photonics*, 8174: 81740- 81741.
41. Ait-Amir, B., Pougnet, P. and El Hami, A. (2020). Meta-model development in embedded mechatronic systems 2. *International Society for Technology in Education*: pp. 157-187.