Engineering Journal issn: 0125-8281

Article

Investigation of Parameters That Affect the Acquired Near Infrared Diffuse Reflected Signals in Non-Destructive Soluble Solids Content Prediction

Fan Wei Hong[1,a] **and Kim Seng Chia**[1,2,b,*]

1 Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Malaysia
2 Microelectronics and Nanotechnology-Shamsuddin Research Center, Institute for Integrated Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia
E-mail: [a]ge180103@siswa.uthm.edu.my, [b]kschia@uthm.edu.my (Corresponding author)

Abstract. Near infrared spectroscopy is a susceptible technique which can be affected by various factors including the surface of samples. According to the Lambertian reflection, the uneven and matte surface of fruits will provide Lambertian light or diffuse reflectance where the light enters the sample tissues and that uniformly reflects out in all orientations. Bunch of researches were carried out using near infrared diffuse reflection mode in non-destructive soluble solids content (SSC) prediction whereas fewer of them studying about the geometrical effects of uneven surface of samples. Thus, this study aims to investigate the parameters that affect the near infrared diffuse reflection signals in non-destructive SSC prediction using intact pineapples. The relationship among the reflectance intensity, measurement positions, and the SSC value was studied. Next, three independent artificial neural networks were separately trained to investigate the geometrical effects on three different measurement positions. Results show that the concave surface of top and bottom parts of pineapples would affect the reflectance of light and consequently deteriorate the predictive model performance. The predictive model of middle part of pineapples achieved the best performance, i.e. root mean square error of prediction (RMSEP) and correlation coefficient of prediction (Rp) of 1.2104 °Brix and 0.7301 respectively.

Keywords: Near infrared, diffuse reflectance, geometrical effects, soluble solids content.

ENGINEERING JOURNAL Volume 24 Issue 6
Received 24 October 2019
Accepted 31 August 2020
Published 30 November 2020
Online at https://engj.org/
DOI:10.4186/ej.2020.24.6.79

DOI:10.4186/ej.2020.24.6.79

1. Introduction

Since the pioneer work of determining the moisture content in agricultural products [1], near-infrared spectroscopy (NIRS), a rapid and non-invasive technique in qualitative and quantitative analysis, has been widely implemented and applied in various fields i.e. agriculture [2]–[8], medical [9], [10], drug [11], and textiles [12]. NIRS is using the near infrared (NIR) region, a part of electromagnetic radiation, which will be partially absorbed by molecular bond consisted hydrogen atom e.g. C-H and O-H to provide the molecular overtone and vibrations information [13]. The near infrared spectra intertwined with the specific information about the composition of examined samples. Using machine learning technology, the hidden information can be extracted and used in discrimination or quality and composition prediction of samples. The initial step in applying NIRS for specific discrimination and prediction is using one of the three spectral data acquisition modes (i.e. transmittance, reflectance, and interactance) [14]. Different samples may contain different compositions, density, volume, and geometrical effects. Thus, a suitable sensing method should be characterized and implemented according to the sample properties.

A typical NIRS set-up contains a NIR light source (e.g. halogen lamp or tungsten lamp), a NIR detector or sensing element, and optical accessories (e.g. slit, optic lens, and grating). The NIR light source plays a vital role in providing NIR energy towards on a sample. Conventionally, due to the ability in covering wide NIR region (780 to 2500nm), halogen tungsten lamp was reported in most studies. However, this NIR energy source has some unavoidable shortcomings which might degrade a NIRS analysis, particularly, it produces unnecessary heat to the sensing environment. This change of environmental temperature will indirectly affect the sample temperature. Consequently, unstable temperature on samples will affect the quality of acquired NIR data. Additionally, in contrast with NIRS rapid and fast characteristics, halogen tungsten lamp needs to be pre–heated before NIR data acquisition. This could slow down the scanning process. For instance, a handheld non-invasive detector mounted with micro-halogen lamp was built to measure the soluble solids content (SSC) (also known as Brix) of fruits [15]. The device was needed to be powered on for at least around 10 minutes before it can proceed to a scanning process. Thus, light emitting diode (LEDs) that has less heat dissipated, insensitive to vibration, and fast operational time to reach maximum light intensity characteristics was believed to be a new type of NIR energy source in NIRS applications [16].

Specifically, NIR diffuse reflection has revealed its potential as a non-destructive technique for rapid analysis in different applications e.g. agriculture [17]–[20], traditional medicine [21], and medical [22]. According to the Lambertian reflection law, diffuse reflection is the reflection light from a matte surface that are scattered from all angles except the specular reflection. Agricultural products especially pineapples that have matte surface compared with apple, pear, and stonefruit. Consequently, an investigation on the NIR measurement positions of samples with matte surface could lead to a better understanding of diffuse reflectance.

For samples with smaller size (e.g. tomato [23], pear [24], apple [25]), the NIR data acquisition process that was performed around the equator showed satisfied prediction results. A research focused on the effects of measurement positions in determining the SSC of apples using on-line NIR system was carried out [26]. Seven models were developed including six local models for six fixed positions, and one global model for combination of all positions. The local model that was built with NIR data collected from equator gave the best prediction accuracy. Similar result was reported by [27] using the similar samples, i.e. apples. However, no similar study has been conducted using fruits with medium size e.g. pineapples.

NIRS can be influenced by various factors e.g. temperature [28], sample biological variability (e.g. cultivar, season, and origin) [29]–[32], sample position detection [27], distance between the sensing instrument and samples [33], angle of illumination and detection (geometry set up) [34], and surface condition of samples [35]. Abundant studies of NIR diffuse reflectance were carried out in non-destructively SSC determination, however, lack of studies concerns about the underlying mechanism of NIR diffuse reflectance. This could limit the understanding of the mechanism and the potential of NIRS. A better understanding will improve the accuracy of non-destructive internal quality measurement in terms of better diffuse reflectance acquisition design to produce a NIR signal with higher signal-to-noise ratio for the internal quality determination. A high accuracy of internal quality prediction will help the farmer in post-harvesting activities e.g. to reduce the unnecessary fruit wastes that could indirectly improve the global fruit security. Therefore, in this study, the parameters that affect the NIR diffuse reflectance in SSC prediction were studied using a sensing instrument that comprised of five NIR LEDs with different wavelengths i.e. 780, 851, 870, 910, and 940 nm.

2. Materials and Methods

2.1. Sample Preparation and NIR Data Acquisition

A total of 75 fresh harvested pineapples were purchased from local market in Batu Pahat, Johor, Malaysia. All pineapples were separated into six batches and collected on different days i.e. 11th, 12nd, 18th, 19th, 24th, and 25th April 2019, respectively. In order to ensure that all the pineapples were stood alone from each other pineapple attributes, each pineapple was washed carefully to remove grimy substances and viscous liquid on the surface of the fruits. This precaution step is to protect the reliability of the data collected. After that,

all pineapples were placed under room temperature for around 20 minutes to ensure their surface was naturally dried before data collection. The soluble solids content (SSC) of pineapples is different from the top to the bottom parts of the fruits. Thus, three data were acquired from a single pineapple i.e. from the top, middle, and bottom of a fruit. Eventually, 225 data were collected (top: 75, middle: 75, and bottom: 75).

A portable sensing instrument (built in with five light emitting diodes (LEDs) with different wavelengths i.e. 780nm, 851nm, 870nm, 910nm, and 940nm, and a photodiode detector) was used to collect the diffuse reflected near infrared (NIR) data non-destructively from the pineapples surface. These wavelengths are at the vicinity of informative wavelengths of the third overtone of OH stretching (780nm) [36]; the third combination overtone of sugar OH stretching at 840nm (851nm) [37]; the strong absorption of water and oxygen (870nm) [38] that was used to predict the SSC of citrus [39]; the third overtone of CH stretching (910nm) [40]; and the strong water absorption band for the second and third overtone of the OH stretching (940nm) [41]. The illumination angle of LED is positioned to 45°. Detail explanation about the working principle of the sensing instrument was reported by [42]. The sensing instrument was undergone calibration process (i.e. dark and white calibration) to calibrate the acquired NIR light intensity of each wavelength before the scanning process. White calibration was carried out with Labsphere Certified Reflectance standard which is able to provide up to 99% reflectance values for 250 to 2500nm. Dark and white reference values were recorded before the data acquisition. To reduce the impact of light and noise, the dark and white references values were obtained under the same condition [43]. Fig. 1 illustrates the schematic diagram of the scanning process. The sensing instrument was directly contacted with the surface of the pineapple. The scanning of each point was repeated three times, and an average of these three successive scans was recorded and used as the inputs of ANN. All the collected data were stored in a computer and processed via software MATLAB R2019a.

2.2. Soluble Solids Content Acquisition

The reference SSC was measured immediately once the NIR data acquisition of pineapples was completed. The flesh along with the peel that was at the data acquisition point was taken out using a stainless steel corer with a dimension of 50mm depth and 20mm diameter. The length of each specimen was limited to 20mm while the surplus flesh was removed. Then, the specimen with peel was crushed using a stainless steel crusher to extract its juice into a beaker. The beaker was shaken gently to ensure the equilibrium of samples. Eventually, few drops of the pineapple juices from the beaker were placed on a digital refractometer (PAL-1, Atago, Tokyo, Japan) to measure the SSC values. The SSC value was expressed into °Brix.

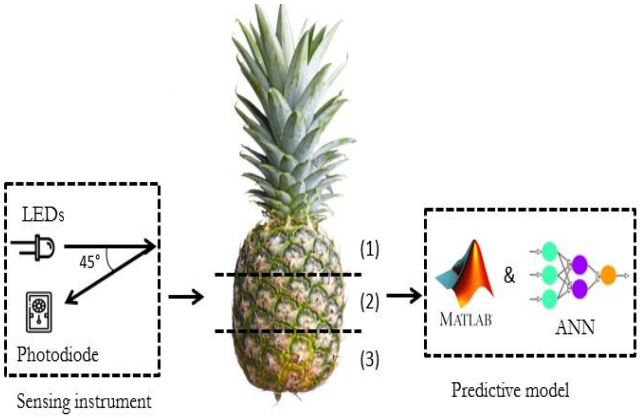

Fig. 1. The scanning position is divided into three different parts along a pineapple: (1) the top part, (2) the middle part, and (3) the bottom part.

2.3. Data Pre-processing

The calibrated diffuse reflectance (Ic) was calculated from the raw NIR data by subtracting the raw data (I_{led}) with dark reference value (I_d), and then that divided by the subtracted value of the white reference (I_w) and I_d as shown in Eq. (1) [44].

$$I_c = \frac{I_{led} - I_d}{I_w - I_d} \quad (1)$$

The calibrated diffuse reflectance was then pre-processed using the normalization and standard normal variate (SNV). The normalization was used to reduce the data redundancy and improve data integrity. SNV was used to reduce the effect of scattering and to correct the baseline shift [45]. Normalization was carried out by dividing the Ic of a particular wavelength with the summation of the Ic of the five different wavelengths at that particular spectrum. SNV was performed by subtracting the Ic with the mean of the spectrum and then divided by the standard deviation of the spectrum.

2.4. Model Development

In this study, three layers (i.e. input, hidden, and output layers) artificial neural network (ANN) model was developed for each measurement position i.e. top, middle, and bottom, respectively. The ANN was trained using Levenberg-Marquardt training algorithm via MATLAB R2019a. Holdout validation was applied i.e. dataset was randomly split into training (70%) and testing (30%) sets that were used to build the calibration model and validate the calibrated model, respectively.

First, the input (i.e. calibrated diffuse reflectance, Ic) and target (i.e. measured soluble solids content (SSC)) were mapped in the range between -1 and +1. Second, holdout validation was applied to determine the optimal calibrated model. The calibrated model was validated with the testing set. The hidden neuron was varied from one to 10. Third, the performance of ANN was evaluated based on the root mean square error (RMSE)

and correlation coefficient (R). The optimal model should achieve the lowest RMSE and the highest R [46]. The RMSE was calculated using Eq. (2), where, n is the number of samples. The best hidden neuron was selected based on the lowest root mean square error of prediction.

$$RMSE=\sqrt{\frac{\Sigma(SSC_{predicted}-SSC_{actual})^2}{n}} \quad (2)$$

Three parameters that affect the near infrared diffuse reflected signals were studied i.e. distribution of light intensity on different parts, relationship between the light intensity and soluble solids content (SSC), and the effect of measurement position. The distribution of light intensity was analyzed through boxplot analysis. Subsequently, the relationship between the light intensity and SSC was investigated using a scatter plot. The effects of measurement positions were investigated by developing three calibration models for different measurement positions, respectively.

3. Results and Discussion

3.1. White and Dark Calibration

Table 1 shows the acquired NIR light intensity values of white and dark calibrations and their calculated diffuse reflectance for five different wavelengths. The calculated diffuse reflectance revealed that the NIR light intensity of each wavelength was maximized without saturation i.e. more than 1024 for 10-bit analog-to-digital conversion.

Table 1. The calibration values of the five wavelengths for white and dark calibrations.

Wavelength (nm)	White Calibration (a.u.)	Dark Calibration (a.u.)	Diffuse Reflectance (%)
780	1001	19	100
851	1000	33	100
870	1001	28	100
910	999	41	100
940	1001	27	100

3.2. Soluble Solids Content (SSC) Reference

Table 2 summarizes the descriptive statistics of acquired soluble solids content (SSC). The SSC values of the top, middle, and bottom parts were normally distributed with mean values of 11.76, 13.25, and 14.23 °Brix and standard deviations of 2.09, 1.93, and 1.94 °Brix, respectively. The ranges of SSC for the three different parts i.e. top, middle, and bottom were 6.4-16.1, 7.3-16.9, and 7.8-19.9 °Brix, respectively. The results indicate that the acquired SSC covered a range from 6.4 to 19.9 °Brix. This range is considered sufficient in developing a good predictive model [27]. Next, the mean value of SSC shows that the top part of the pineapples has the lowest SSC while the bottom part has the highest SSC. This could be due to the ripening process that begins from the bottom to the top parts of pineapples. The descriptive statistics of the measured SSC is in-line with the distribution of internal quality of pineapples study that reported by [47].

Table 2. Descriptive statistics of the measured SSC for three different parts of pineapples

Pineapple Parts	Number of samples	Min (°Brix)	Max (°Brix)	Mean (°Brix)	SD (°Brix)
Top	75	6.4	16.1	11.76	2.09
Middle	75	7.3	16.9	13.25	1.93
Bottom	75	7.8	19.9	14.23	1.94
Total	225	6.4	19.9	13.08	2.24

3.3. Overview of NIR Data and Pattern Recognition

Table 3 indicates the descriptive statistics of the 16 patterns of the acquired NIR light. Each part has 75 NIR samples. The result indicates that the top four identical patterns had occupied around 70% of the acquired NIR light in the top, middle, and bottom parts, respectively. For instance, the number of samples that involved in the first four ranking patterns contribute 76% for the middle part; whereas that contributes 72% in the top and bottom parts. This implies that the acquired data have the higher consistency on the middle part compared with the top and the bottom parts. This might be due to less geometrical effects on the middle part of pineapples that have relatively flat surface compared with the top and bottom parts.

Table 3. Descriptive statistics of the first four ranking patterns

Pattern	Top		Middle		Bottom	
	Sample	%	Sample	%	Sample	%
1	21	28	23	30.67	16	21.33
2	14	18.67	22	29.33	18	24
3	8	14.67	7	9.33	18	24
4	11	10.67	5	6.67	2	2.67
Others	21	28	18	24	21	28
Total	75	100	75	100	75	100

Each wavelength was compared with the following wavelength in a sequence from 780nm to 940nm to determine its particular pattern. After pre-processed the raw diffuse reflectance data from three different parts i.e. the top, middle, and bottom of pineapples, and compared the acquired NIR light intensity difference among the five different wavelengths, 16 patterns were observed. Fig. 2 illustrates the top four ranking patterns that were obtained. For instance, pattern 1 (dot line) was defined where the light intensity of 780nm is lower than 851nm and then 851nm lower than 870nm.

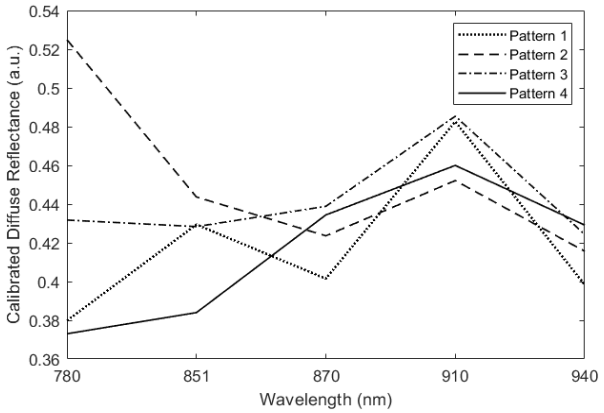

Fig. 2. The top four ranking patterns: Dot line represents Pattern 1, Dash line represents Pattern 2, Dot-dash line represents Pattern 3, and Solid line represents Pattern 4.

3.4. Investigated Parameters

3.4.1. Reflected NIR Intensity in Different Parts

Fig. 3 illustrates the boxplot of five near infrared (NIR) wavelengths light intensity values on the three different parts of pineapples (i.e. top, middle, and bottom). This technique (boxplot) can identify potential outliers on each wavelength with dependent y-variable (i.e. diffuse reflectance) which fall outside the 99.9% coverage of a normally distributed data [5]. Five independent wavelengths under different parts reveal insignificant interquartile range (length of the box) and also whisker (spread from two ends of box until its reach sample maximum and minimum value). This phenomenon reflects the irregularly soluble solids content (SSC) distribution on different parts of pineapples.

The whisker helps in distinguishing the outlier data from the normal distributed dataset. For instance, wavelength 780nm in the boxplot of top and middle part does not discover any outlier data (i.e. red plus symbol), whereas there were five outlier data at the bottom part. In total, 46 outliers were recorded from the three different parts, i.e. top: 13, middle: 12, bottom: 21. Higher number of outlier data in the top and bottom parts might be caused by the concave surface of pineapples which will alter the ways that NIR light was reflected from the sample tissues [26]. Besides, the number of outlier data suggests that the consistency of the middle part was the highest because it had the lowest number of outlier data. This is in-line with the first four patterns that discussed in Section 3.3, i.e. the acquired NIR light from the middle part of pineapples had the highest consistency.

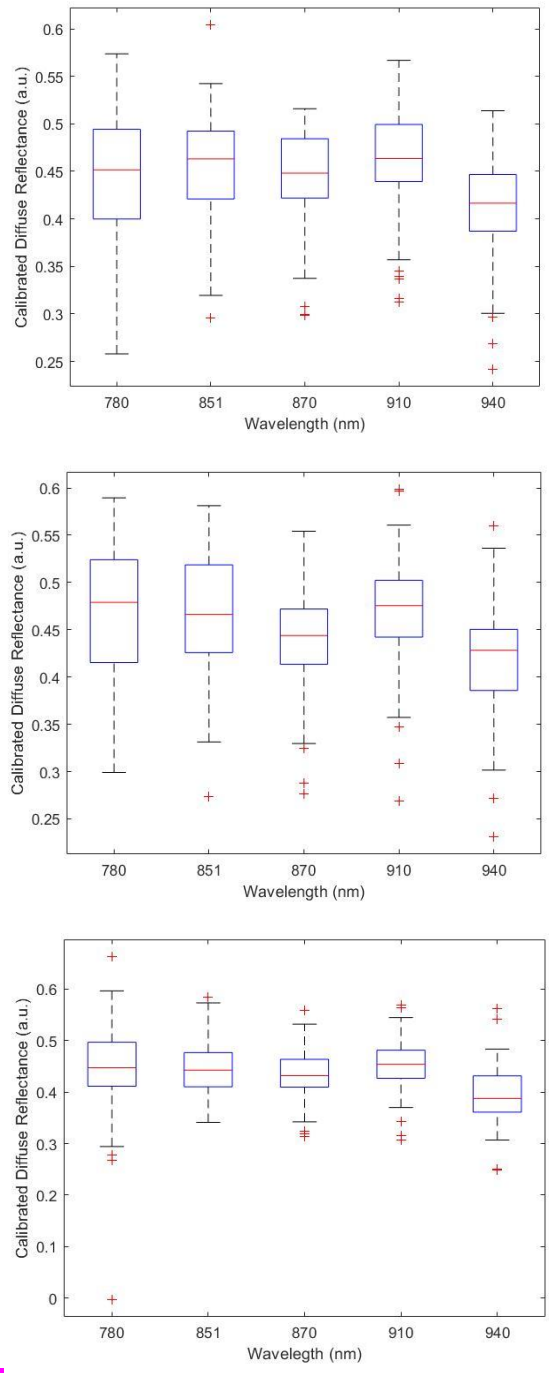

Fig. 3. Boxplot: (a) Top, (b) Middle, (c) Bottom (total potential outliers: top: 13, middle: 12, bottom: 21).

3.4.2. The Relationship between the Light Intensity and Soluble Solids Content

Fig. 4 illustrates 15 scatter plots of soluble solids content (SSC) versus the calibrated NIR diffuse reflectance on different parts of pineapples. The result indicates that the correlation coefficient (R) values are relatively low where the highest and lowest values were only 0.2224 for 940nm from the middle part and 0.0063 for 910nm from the top part, respectively. R values of 940nm and 910nm from middle part gave higher values than that from other parts. This might be due to higher consistency of the acquired NIR from the middle.

DOI:10.4186/ej.2020.24.6.79

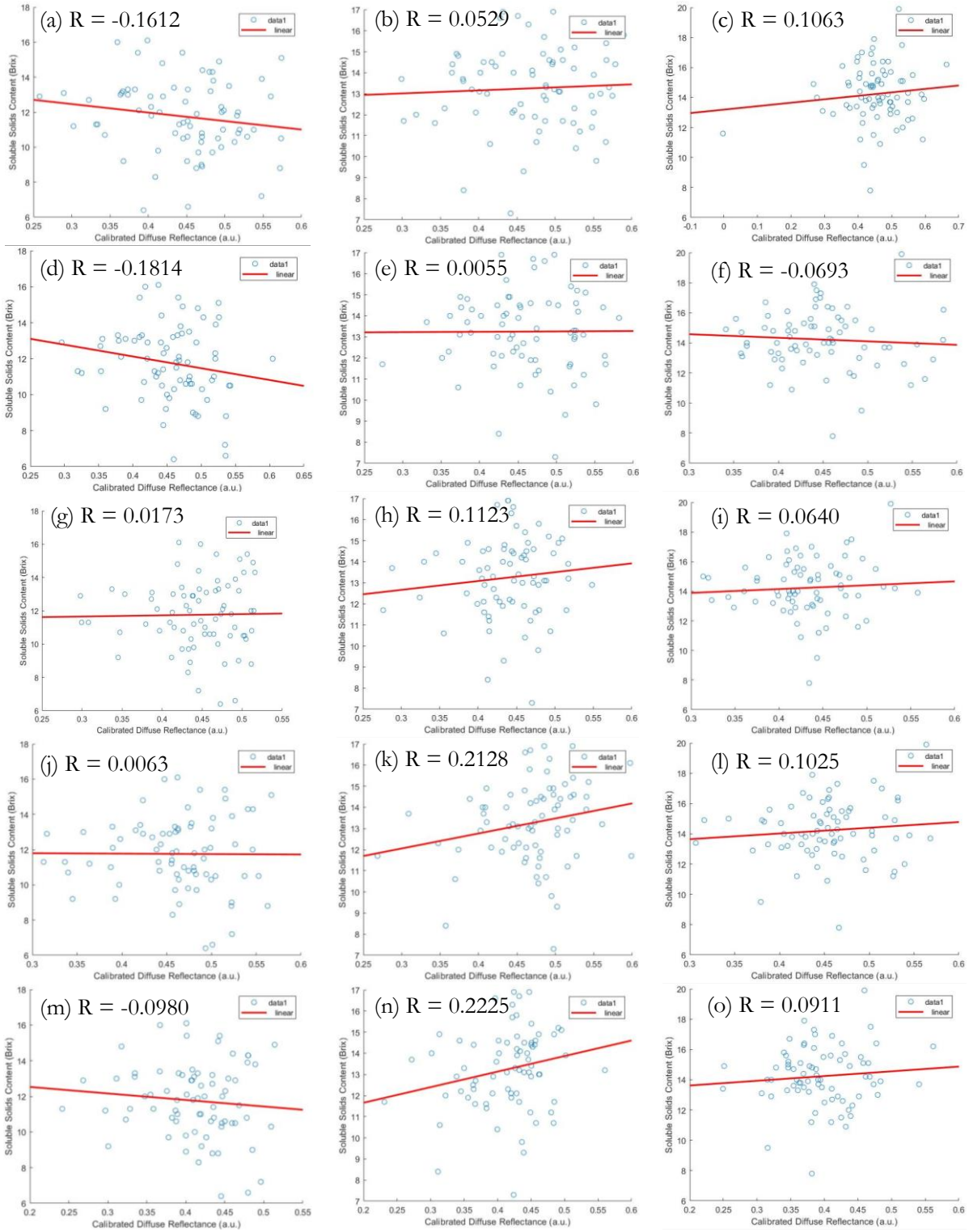

Fig 4. The correlation between SSC and diffuse reflectance of each wavelength: (a), (b), and (c) are the top, middle, and bottom for 780nm; (d), (e), (f) are that for 851nm; (g), (h), (i) are that for 870nm; (j), (k), (l) are that for 910nm; and (m), (n), (o) are that for 940nm.

3.4.3. Effect of Pre-Processing

Nine artificial neural network (ANN) models i.e. top, middle, and bottom were trained with and without pre-processing methods. The prediction performance of different independent ANN models was validated using respective testing data. For instance, the top ANN model was trained using 70% of the data that acquired from the top part, and then evaluated using the rest 30% of the data. The results of training and testing in terms of root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), correlation coefficient of calibration/training (Rc), and correlation

coefficient of prediction/testing (Rp) are tabulated in Table 4.

Table 4 shows that the performance of ANN with the SNV pre-processing method achieved better performance for the top and middle parts, followed by that without pre-processing and that with normalization. The performance of ANN without pre-processing was better than the normalization might be due to the pre-processing technique removed interested signals, thus, degraded the performance of ANN. Meanwhile, the performance of bottom ANN model contrasts with top and middle model where the normalization achieved best performance, followed by that without pre-processing and that with the SNV. However, the overall calibration and prediction performance of the bottom model was inferior to the top and middle models.

3.4.4. Prediction Performance of ANN Model

Fig. 5 illustrates the relationship between the predicted SSC and measured SSC for the training and testing by three calibrated models with SNV i.e. top, middle, and bottom, respectively. The RMSEC and Rc of three different measurement positions for SNV model i.e. top, middle, and bottom are 1.4921 °Brix, 0.7285; 1.0998 °Brix, 0.8394; and 1.8419 °Brix, 0.4799 respectively. The middle part revealed the best training performance with lowest RMSEC and highest Rc. The similar geometrical effect of the concave surface on the top and bottom parts of pineapples may the reason that led to an unsatisfied training performance. This observation is in-lined with the previous study [26] where the training performance is similar at the concave surface of calyx and stem end of the apples. The non-uniform surface of pineapples limits the penetration of NIR light to obtain sufficient absorbance. For samples liked durian, this problem can be minimised by destructively removing some of the peel to provide a flat area for scanning [50]. Since the surface of pineapples sticking with its flesh, this approach is unsuitable to be applied to improve the scanning method.

Table 4. Calibration and prediction performance of ANN for three parts i.e. calibrated diffuse reflectance without pre-processing (no applied), that with normalization, and that with standard normal variate (SNV)

Pineapple parts	Different Pre-processing method	Hidden Neurons	Calibration		Prediction	
			RMSEC	Rc	RMSEP	Rp
Top	No applied	2	1.8162	0.5520	1.7012	0.5206
	Normalization	2	1.9221	0.4704	1.6451	0.3343
	SNV	3	1.4921	0.7285	1.4866	0.5509
Middle	No applied	2	1.7453	0.5062	1.3719	0.5285
	Normalization	2	1.7096	0.5351	1.4792	0.4573
	SNV	5	1.0998	0.8394	1.2104	0.7301
Bottom	No applied	1	2.0258	0.2627	1.3333	0.3932
	Normalization	1	2.0351	0.2458	1.2439	0.4866
	SNV	3	1.8419	0.4799	1.3412	0.3139

Next, the prediction accuracy of the calibrated models was evaluated based on RMSEP and Rp. The RMSEP and Rp for three SNV calibrated models i.e. top, middle, and bottom were 1.4866 °Brix; 0.5509, 1.2104 °Brix; 0.7301, and 1.3412 °Brix; 0.3139 respectively. In previous study [51], non-invasive prediction of soluble solids content (SSC) of pineapples using wavelength ranged from 668 to 968nm achieved RMSEP between 0.82 to 1.01 °Brix. The RMSEP values in this study is relatively high. This may be caused by the lack of wavelength range covered and subsequently lost the small portion of information. This is in-line with previous study that used a simplified LED design with four NIR wavelengths from 630 to 890nm in rapid ripeness evaluation of white grape [37].

Fig. 5 illustrates that the best prediction was achieved by the ANN model for the middle, followed by the top, and the bottom parts. The best number of hidden neurons in the middle model was five which is same with the previous research [49] where ANN with five hidden neurons gave 68.89% in correctly classification of internal quality of pineapples. Next, the optimum model was developed using the acquired NIR data on the middle/equator part of pineapples. This selection was based on the prediction accuracy that illustrated in Fig. 5.

Next, previous study reported the similar trend in investigating the effect of measurement positions on apples where the equator model gave the best performance with RMSEP = 0.447°Brix and Rp= 0.970 [27]. Similar researches that acquired NIR data from the equator part performed well in prediction model e.g. tomato (r^2= 0.9718, RMSEP = 0.1707) [23], and pear (r^2= 0.83, RMSEP = 0.57) [24]. Even though the prediction performance of this present study appears to be low with RMSEP = 1.2104 °Brix; Rp = 0.7301, this accuracy is considered comparable with that reported in previous study in SSC prediction of loquats i.e. RMSEP= 1.180, Rp = 0.717 [52].

DOI:10.4186/ej.2020.24.6.79

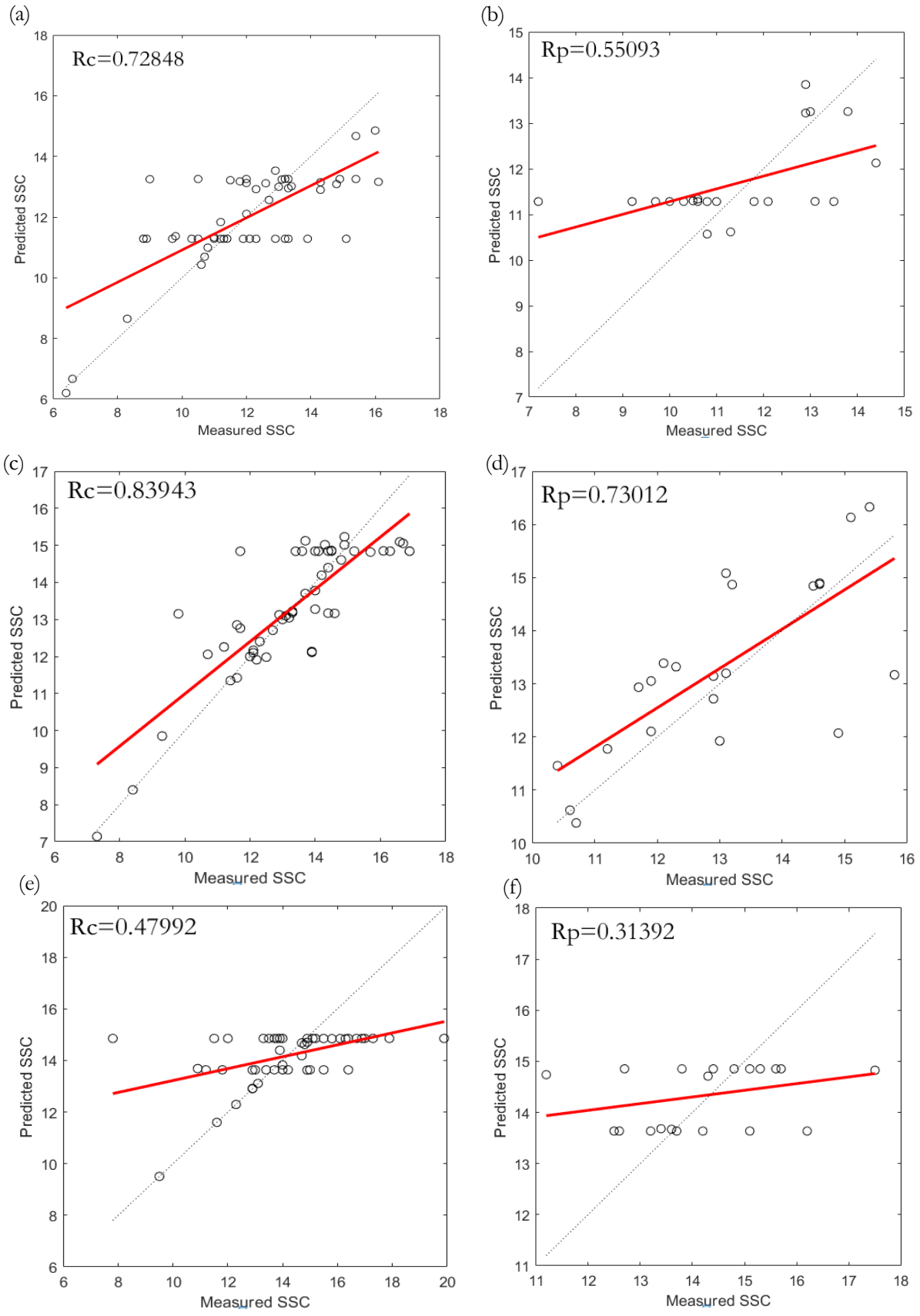

Fig 5. Correlation plot of SSC prediction with SNV calibrated model: (a) R_c for the top, (b) R_p for the top; (c) R_c for the middle, (d) R_p for the middle; (e) R_c for the bottom, (f) R_p for the bottom parts of pineapple.

4. Conclusion and Recommendation

Three parameters that affect the near infrared diffuse reflected signal in non-destructive soluble solids content (SSC) prediction for intact pineapples were studied and evaluated using artificial neural network (ANN) i.e. effect of light intensity of light emitting diode (LED), the relationship between SSC and near infrared light intensity, and measurement positions. Based on the boxplot analysis, the distribution light intensity of different wavelengths on different parts of pineapples showed that the top and bottom parts of pineapples had higher number of outliers than the middle part. This implied that the effects of the concave surface of pineapples which affects the consistency of the light entering and reflecting from the sample tissues. Next, the scatter plot of the middle part of pineapples in investigating the relationship between light intensity and SSC showed that the near infrared light with wavelengths of 870, 910, and 940nm provided higher consistency results than that at the top and bottom parts. Furthermore, the relationship between the measurement positions and prediction accuracy was established using ANN models with and without preprocessing of normalization and standard normal variate (SNV). Results indicate that the calibrated model with SNV on the middle part of pineapples gave the best prediction performance i.e. RMSEP = 1.2104 °Brix; Rp = 0.7301. These suggest that the best SSC prediction performance can be obtained by avoiding the prediction on the top and bottom parts of pineapples. Future research should study alternatives to minimize the scatter effects of NIR spectrum, and different types of machine learning algorithms that may potentially improve the prediction accuracy of the NIRS.

Acknowledgement

The author acknowledges the funding from Geran Penyelidikan Pascasiswazah Fasa 2/2018 (GPPS) (H312) under Research Management Centre (RMC) in Universiti Tun Hussein Onn Malaysia (UTHM) in completing this study.

References

[1] N. Karl H, "Direct Spectrophotometric Determination of Moisture Content of Grain and Seeds," *J. Near Infrared Spectrosc.*, vol. 4, no. 1, pp. 23–30, 1996.

[2] L. C. Carvalho et al., "Using Intact Nuts and Near Infrared Spectroscopy to Classify Macadamia Cultivars," *Food Anal. Methods*, vol. 11, no. 7, pp. 1857–1866, 2018.

[3] H. Chen, C. Tan, Z. Lin, and T. Wu, "Classification and quantitation of milk powder by near-infrared spectroscopy and mutual information-based variable selection and partial least squares," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 189, pp. 183–189, 2018.

[4] C. Li, H. Guo, B. Zong, P. He, F. Fan, and S. Gong, "Rapid and non-destructive discrimination of special-grade flat green tea using Near-infrared spectroscopy," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 206, pp. 254–262, 2019.

[5] J. C. Hashimoto et al., "Quality control of commercial cocoa beans (Theobroma cacao L.) by near-infrared spectroscopy," *Food Anal. Methods*, vol. 11, no. 5, pp. 1510–1517, 2018.

[6] F. Caro, M. Constantino, I. Martins, and A. Weintraub, "PLS, iPLS, GA-PLS models for soluble solids content, pH and acidity determination in intact dovyalis fruit using near-infrared spectroscopy," *For. Sci.*, vol. 49, no. 5, pp. 738–751, 2018.

[7] Z. Li *et al.*, "Identification of oil, sugar and crude fiber during tobacco (Nicotiana tabacum L.) seed development based on near infrared spectroscopy," *Biomass and Bioenergy*, vol. 111, no. August 2017, pp. 39–45, 2018.

[8] N. Nakawajana, J. Posom, and J. Paeoui, "Prediction of higher heating value, lower heating value and ash content of rice husk using FT-NIR spectroscopy," *Eng. J.*, vol. 22, no. 5, pp. 45–56, 2018.

[9] T. Kishino et al., "Predictive evaluation of pharmaceutical properties of ulinastatin-containing vaginal suppositories as a hospital preparation by near-infrared spectroscopy," *Chem. Pharm. Bull.*, vol. 66, no. 6, pp. 589–595, 2018.

[10] P. Kadamati et al., "Near-infrared spectroscopy muscle oximetry of patients with postural orthostatic tachycardia syndrome," *J. Innov. Opt. Health Sci.*, vol. 11, no. 05, p. 1850026, 2018.

[11] C. mei Liu, Y. Han, S. geng Min, W. Jia, X. Meng, and P. pei Liu, "Rapid qualitative and quantitative analysis of methamphetamine, ketamine, heroin, and cocaine by near-infrared spectroscopy," *Forensic Sci. Int.*, vol. 290, pp. 162–168, 2018.

[12] H. Yan and H. W. Siesler, "Identification of textiles by handheld near infrared spectroscopy: Protecting customers against product counterfeiting," *J. Near Infrared Spectrosc.*, vol. 26, no. 5, pp. 1–11, 2018.

[13] V. E. Dabkiewicz, S. de Mello Pereira Abrantes, and R. J. Cassella, "Development of a non-destructive method for determining protein nitrogen in a yellow fever vaccine by near infrared spectroscopy and multivariate calibration," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 201, no. 2017, pp. 170–177, 2018.

[14] B. M. Nicolaï et al., "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review," *Postharvest Biol. Technol.*, vol. 46, no. 2, pp. 99–118, 2007.

[15] W. Guo, W. Li, B. Yang, Z. Z. Zhu, D. Liu, and X. Zhu, "A novel noninvasive and cost-effective handheld detector on soluble solids content of fruits," *J. Food Eng.*, vol. 257, pp. 1–9, 2019.

[16] J. Malinen and M. Ka, "LED-based NIR spectrometer module for hand-held and process analyser applications," vol. 51, pp. 220–224, 1998.

[17] A. K. Hazarika et al., "Quality assessment of fresh tea leaves by estimating total polyphenols using near infrared spectroscopy," *J. Food Sci. Technol.*, vol. 55, no. 12, pp. 4867–4876, 2018.

[18] D. Eisenstecken, B. Stürz, P. Robatscher, L. Lozano, A. Zanella, and M. Oberhuber, "The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations," *Postharvest Biol. Technol.*, vol. 147, pp. 123–131, 2019.

[19] Y. Cui et al., "Identification of maize seed varieties based on near infrared reflectance spectroscopy and chemometrics," *Int. J. Agric. Biol. Eng.*, vol. 11, no. 2, pp. 177–183, 2018.

[20] M. Lopo, C. A. Teixeira dos Santos, R. N. M. J. Páscoa, A. R. Graça, and J. A. Lopes, "Near infrared spectroscopy as a tool for intensive mapping of vineyards soil," *Precis. Agric.*, vol. 19, no. 3, pp. 445–462, 2018.

[21] P. Liu, J. Wang, Q. Li, J. Gao, X. Tan, and X. Bian, "Rapid identification and quantification of Panax notoginseng with its adulterants by near infrared spectroscopy combined with chemometrics," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 206, pp. 23–30, 2019.

[22] Y. Shimokawa, E. Hayakawa, K. Takahashi, K. Okai, Y. Hattori, and M. Otsuka, "Pharmaceutical formulation analysis of gelatin-based soft capsule film sheets using near-infrared spectroscopy," *J. Drug Deliv. Sci. Technol.*, vol. 48, no. April, pp. 174–182, 2018.

[23] S. Jun, L. Yating, W. Xiaohong, D. Chunxia, and C. Yong, "SSC prediction of cherry tomatoes based on IRIV-CS-SVR model and near infrared reflectance spectroscopy," *J. Food Process Eng.*, vol. 41, no. 8, p. e12884, 2018.

[24] Y. Liu, W. Liu, X. Sun, R. Gao, Y. Pan, and A. Ouyang, "Potable NIR spectroscopy predicting soluble solids content of pears based on LEDs," *J. Phys. Conf. Ser.*, vol. 277, pp. 1–11, 2011.

[25] X. Tian, J. Li, Q. Wang, S. Fan, and W. Huang, "A bi-layer model for nondestructive prediction of soluble solids content in apple based on reflectance spectra and peel pigments," *Food Chem.*, 2017.

[26] X. Xu, H. Xu, L. Xie, and Y. Ying, "Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system," *J. Food Meas. Charact.*, vol. 13, no. 1, pp. 506–512, 2018.

[27] S. Fan, B. Zhang, J. Li, W. Huang, and C. Wang, "Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple," *Biosyst. Eng.*, vol. 143, pp. 9–19, 2016.

[28] D. Cheng, W. Cai, and X. Shao, "Understanding the interaction between oligopeptide and water in aqueous solution using temperature-dependent near-infrared spectroscopy," *Appl. Spectrosc.*, vol. 72, no. 9, pp. 1354–1361, 2018.

[29] S. Shrestha, L. Deleuran, and R. Gislum, "Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics," *J. Spectr. Imaging*, vol. 5, no. 1, pp. 1–9, 2016.

[30] F. M. G. Ramalho, J. M. Andrade, and P. R. G. Hein, "Rapid discrimination of wood species from native forest and plantations using near infrared spectroscopy," *For. Syst.*, vol. 27, no. 2, 2018.

[31] X. Li, J. Huang, Y. Xiong, J. Zhou, X. Tan, and B. Zhang, "Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy," *Comput. Electron. Agric.*, vol. 155, pp. 23–31, 2018.

[32] S. Munera, J. M. Amigo, J. Blasco, S. Cubero, P. Talens, and N. Aleixos, "Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging," *J. Food Eng.*, vol. 214, pp. 29–39, 2017.

[33] Z. Tan, J. Xie, J. Chen, C. Ding, and K. Monte, "Optimization of detection device geometry for NIR spectroscopy using a three-layered model of stone fruit," *Opt. Rev.*, vol. 23, no. 5, pp. 784–790, 2016.

[34] B. Zhang, D. Dai, J. Huang, J. Zhou, Q. Gui, and F. Dai, "Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review," *Crit. Rev. Food Sci. Nutr.*, vol. 58, no. 12, pp. 2099–2118, 2018.

[35] A. Phuphaphud, K. Saengprachatanarug, J. Posom, S. Wongpichet, K. Maraphum, and E. Taira, "Effects of waxy types of a sugarcane stalk surface on the spectral characteristics of visible-shortwave near infrared measurement," *Eng. J.*, vol. 23, no. 1, pp. 13–24, 2019.

[36] A. Clément, M. Dorais, and M. Vernon, "Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using visible NIR spectroscopy," *J. Agric. Food Chem.*, vol. 56, no. 21, pp. 9813–9818, 2008.

[37] V. Giovenzana, R. Civelli, R. Beghi, R. Oberti, and R. Guidetti, "Testing of a simplified LED based vis / NIR system for rapid ripeness evaluation of white grape (Vitis vinifera L.) for Franciacorta wine," *Talanta*, vol. 144, pp. 584–591, 2015.

[38] Y. Zhang, L. Zheng, M. Li, X. Deng, and R. Ji, "Predicting apple sugar content based on spectral characteristics of apple tree leaf in different phenological phases," *Comput. Electron. Agric.*, vol. 112, pp. 20–27, 2015.

[39] X. Fu, X. Wang, and X. Rao, "An LED-based spectrally tuneable light source for visible and near-infrared spectroscopy analysis : A case study for

sugar content estimation of citrus," *Biosyst. Eng.*, vol. 163, pp. 87–93, 2017.

[40] A. M. Cavaco et al., "Validation of short wave near infrared calibration models for the quality and ripening of ' Newhall ' orange on tree across years and orchards," *Postharvest Biol. Technol.*, vol. 141, pp. 86–97, 2018.

[41] Y. Liu, X. Sun, H. Zhang, and O. Aiguo, "Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy," *Comput. Electron. Agric.*, vol. 71, pp. 10–14, 2010.

[42] M. N. H. Jam and K. S. Chia, "A five band near-infrared portable sensor in nondestructively predicting the internal quality of pineapples, in *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA) IEEE*, March 2017, pp. 135-138.

[43] M. Sun, D. Zhang, L. Liu, and Z. Wang, "How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method," *Food Chem.*, vol. 218, pp. 413–421, 2017.

[44] I. Kavdir, R. Lu, D. Ariana, and M. Ngouajio, "Visible and near-infrared spectroscopy for nondestructive quality assessment of pickling cucumbers," *Postharvest Biol. Technol.*, vol. 44, no. 2, pp. 165–174, 2007.

[45] X. Lin, J. L. Xu, and D. W. Sun, "Investigation of moisture content uniformity of microwave-vacuum dried mushroom (Agaricus bisporus) by NIR hyperspectral imaging," *Food Sci. Technol.*, vol. 109, pp. 108–117, 2019.

[46] E. Arendse, O. A. Fawole, L. S. Magwaza, H. Nieuwoudt, and U. L. Opara, "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. Technol.*, vol. 138, pp. 91–98, 2018.

[47] E. V. Miller and G. D. Hall, "Distribution of total soluble solids, ascorbic acid, total acid, and bromelin activity in the fruit of the natal pineapple (*Ananas comosus* L . Merr.)," *Plant Physicology*, vol. 28, no. 3, pp. 532–534, 2016.

[48] Z. Genisheva, C. Quintelas, D. P. Mesquita, E. C. Ferreira, J. M. Oliveira, and A. L. Amaral, "New PLS analysis approach to wine volatile compounds characterization by near infrared spectroscopy (NIR)," *Food Chem.*, vol. 246, no. October 2017, pp. 172–178, 2018.

[49] M. Nur, H. Jam, and K. S. Chia, "Investigating the relationship between the reflected near infrared light and the internal quality of pineapples using neural network," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 4, pp. 1389–1394, 2017.

[50] K. Phuangsombut, A. Phuangsombut, A. Talabnark, and A. Terdwongworakul, "Empirical reduction of rind effect on rind and flesh absorbance for evaluation of durian maturity using near infrared spectroscopy," *Postharvest Biol. Technol.*, vol. 142, no. April, pp. 55–59, 2018.

[51] K. S. Chia, H. Abdul Rahim, and R. Abdul Rahim, "Prediction of soluble solids content of pineapple via non-invasive low cost visible and shortwave near infrared spectroscopy and artificial neural network," *Biosyst. Eng.*, vol. 113, no. 2, pp. 158–165, 2012.

[52] X. Fu et al., "Determination of soluble solid content and acidity of loquats based on FT-NIR Spectroscopy," *J. Zhejiang Univ. Sci. B*, vol. 10, no. 2, pp. 120–125, 2009.

DOI:10.4186/ej.2020.24.6.79

Fan Wei Hong was born in Batu Pahat, Johor, Malaysia in 1994. He obtained the BEng in mechatronic and robotic engineering from Universiti Tun Hussein Onn Malaysia (UTHM) at Batu Pahat, Malaysia in June 2018. From 2018 – present, he currently pursuing his MEng in electrical engineering in Faculty of Electrical and Electronic Engineering, UTHM. His research interests are artificial intelligence and near infrared spectroscopic analysis.

Kim Seng Chia was born in Batu Pahat, Johor, Malaysia in 1986. He obtained his PhD (electrical engineering) and BEng (electronic engineering – control & instrumentation) from Universiti Teknologi Malaysia in Feb. 2014 and Sept. 2010, respectively.

He was an assistant professor at Southern University College, Skudai, Malaysia from March 2014 to August 2014. He has served as a lecturer in the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia since September 2014. He is the principal investigator of 8 research projects since 2014. He is the author of more than 40 research articles. His research interests are machine learning, artificial intelligence, real-time embedded system, soft modeling, and near infrared spectroscopic analysis.

Dr Chia was a recipient of the International Invention, Innovation & Technology Exhibition, Malaysia (ITEX) and Seol International Invention Fair Silver awards in 2018.