Article

A Comparison between the Post- and Pre-dispersive Near Infrared Spectroscopy in Non-Destructive Brix Prediction Using Artificial Neural Network

Fan Wei Hong[a], Kim Seng Chia[b,*], and Xien Yin Yap[c]

Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Malaysia
E-mail: [a]hongfanwei0211@gmail.com, [b,*]kschia@uthm.edu.my (Corresponding author), [c]xienyinyap@gmail.com

Abstract. Even though near infrared (NIR) spectroscopy have been implemented in determining the Brix of pineapples, no traceable study compares the effects of different acquisition designs. Thus, this study aims to evaluate the prediction performance of both pre- and post-dispersive NIR sensing devices in non-destructive Brix prediction using artificial neural network (ANN). The pre-dispersive device has five narrowband light emitting diodes (LEDs) with different wavelengths and a photodiode detector, whereas the post-dispersive device has a bifurcated fiber optic, a broadband LED, and a spectral sensor. First, the NIR diffuse reflectance was non-destructively collected using both NIR devices. Then, the collected diffuse reflectance was calibrated with the white and dark references, and then pre-processed using normalization and standard normal variate methods. After that, ANNs were built for both devices using the pre-processed data. Results show both devices are suitable for sample screening application with range error ratio (RER) of more than seven. Nevertheless, the ANN that trained using the post-dispersive device outperformed that trained using the pre-dispersive device with an 8.1% improvement of correlation coefficient of prediction (i.e. from 0.6853 to 0.7408), and a 5.7% improvement of root mean square error of prediction (i.e. from 1.3918 to 1.313°Brix).

Keywords: Near infrared, pre-dispersive, post-dispersive, artificial neural network, pineapple Brix prediction.

1. Introduction

Among various internal qualities of pineapples, the soluble solids content (SSC) (also known as Brix value) is often used as the main quality indicator of pineapples. However, the conventional Brix measurement of pineapple is destructively and time-consuming. Firstly, the pineapple was dissected, and the flesh was squeezed to produce juices. Then, few drops of juices were dropped on a digital refractometer to measure the Brix value. This destructive and time-consuming process is impractical to ensure each exported pineapple complies the minimum Brix requirement i.e. 12 °Brix. Thus, a rapid and non-destructively alternative method should be implemented for pineapple screening based on their Brix value instead of skin colour.

Near infrared (NIR) region is a part of electromagnetic radiation with spectral range from 780 to 2500nm [1]. NIR energy with specific wavelengths will be partially absorbed by specific molecular bonds that consist of hydrogen e.g. C-H and O-H to become mechanical energy of molecular overtone and vibrations [2]. The amount of this energy transformation (also known as absorption) is affected by the chemical composition and the physical properties of an examined sample. Hence, near infrared spectroscopy (NIRS), a rapid and non-invasive sensing technique in qualitative and quantitative analysis using NIR region, has been widely implemented in predicting the SSC of fruit non-destructively e.g. apple [3], pears [4], dovyalis [5], and cherry tomatoes [6]. Additionally, NIRS was reported to be able to predict the SSC of sugarcane [7] and marian plum fruit [8].

NIR acquisition aims to acquire the intensity of NIR spectrum that can be used to predict components of interest of a sample instantly and non-invasively without a sample preparation. Fundamentally, a typical NIRS system comprises of three important components i.e. light source (e.g. a halogen lamp or light emitting diode (LED)), a NIR sensing device, and a sample. There are two NIR acquisition designs i.e. pre-dispersive and post-dispersive that are different in terms of NIR light that is dispersed before or after passing through a sample.

A pre-dispersive NIR acquisition can be implemented using few different NIR LEDs [9] to emit NIR light with different specific wavelengths toward a sample, and then a light sensor to measure the intensity of each wavelength. Alternatively, a broadband light source coupled with bandpass filters can be used to produce NIR light for pre-dispersive NIR acquisition [10]. The pre-dispersive NIR acquisition designed with low costly NIR LEDs with specific different wavelengths and a sensing element is promising in reducing the financial barrier of NIR acquisition [9]. Recently, a pre-dispersive design showed that the collected NIR spectrum was able to achieve RMSEP = 1.2104 °Brix and Rp = 0.7301 in predicting the Brix of pineapple using Artificial Neural Network (ANN) coupled with standard normal variate (SNV) [11]. On the other hand, a post-dispersive NIR acquisition can be conducted using a broadband light source to emit NIR spectrum toward a sample, and then a grating or bandpass filters can be used to disperse or filter the light so that the intensity of the light with specific wavelengths can be measured.

To model NIR signals for qualitative or quantitative prediction applications, various non-linear algorithms (i.e. Artificial Neural Network (ANN), local algorithm (LA), AdaBoost, extreme learning machine (ELM), and support vector machine (SVM)) were studied as that reviewed by M. Zareef et. al. [12]. Among these algorithms, ANN is considered at the most popular non-linear algorithm in NIRS researches. For instance, ANN was used to model NIR signals in predicting dry matter content [13], blood glucose [14], nitrogen content [15], protein content [16], moisture content [17], and zinc oxide content [18].

Nevertheless, NIRS can be influenced by various factors e.g. temperature [19], sample biological variability (cultivar, season, origin) [20]–[23], sample position detection [24], distance between the sensing instrument and samples [25], angle of illumination and detection (geometry set up) [26], and surface condition of samples (geometrical effect) [27]. Notably, the fundamental of both pre- and post-dispersive techniques in eliminating the geometrical effect on the surface of pineapple during data acquisition may have different effects on the Brix prediction performance. A better understanding about the effectiveness of both different designs in the Brix prediction performance may enhance the usefulness of NIR technology. Since there is no traceable study in comparing both post- and pre-dispersive designs, this study aims to evaluate the prediction performance of the pre- and post-dispersive NIR devices in non-destructive Brix prediction for pineapples using artificial neural network (ANN).

2. Materials and Methods

2.1. Sample Preparation and NIR Data Acquisitions

A total of 72 fresh intact pineapples (MD2 species) that harvested from a local pineapple farm in Ladang FIMA Ayer Hitam, Johor, Malaysia were used in this study. All fresh harvested pineapples were directly transported to our laboratory (approximately 25km from the farm) for NIR data acquisition and Brix measurement on the same day. The maturity index of all harvested pineapples were between 0 to 2 that were based on the skin color of pineapples and classified by experienced farmers.

For each pineapple, two "eyes" (also known as fruitlets) around the equator (middle part) of the pineapple were selected, in which, both of them were around 180° apart to each other as illustrated in Fig. 1. Each "eye" should have different Brix value as pineapple is a multiple fruit (also known as collective fruit). After that, the in-house developed pre-dispersive NIR device [9] was used to acquire NIR signals. Each "eye" was scanned five times at different points as illustrated in Fig. 1. The measurement probe was directly contacted on the surface to minimize the effects of ambient light. In short, a total

of ten diffuse reflectance NIR data were non-destructively collected from one intact pineapple. Similar scanning method was applied to get four NIR data at different positions on a Marian plum fruit [8]. The above procedure was repeated using a post-dispersive NIR device before the pineapple was cut for conventional Brix measurement. In total, 720 NIR diffuse reflectance data were collected from 72 pineapples with both devices, respectively.
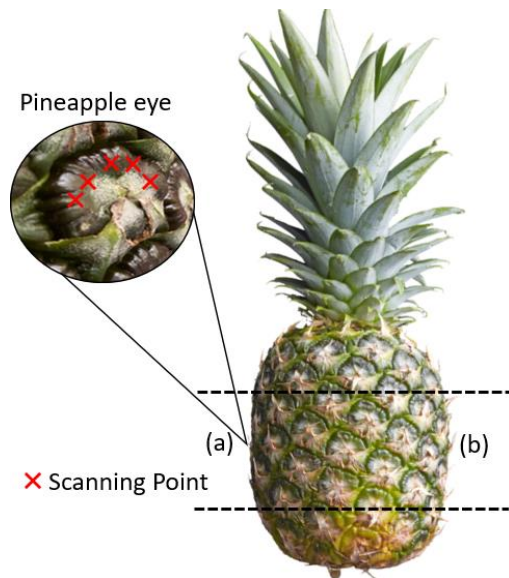

Fig. 1. Five NIR scanning points on a selected "eye" on the middle part of a pineapple, where (a) is allocated 180° with (b).

The pre-dispersive device consisted of an OPT101 monolithic photodiode detector (Texas Instrument) and five light emitting diodes with different wavelengths i.e. 780, 851, 870, 910, and 940nm (Thorlab, Japan). The wavelengths range between 700 and 1000nm have higher radiation and weaker water absorbance to penetrate further in a fruit, and are able to recognize third and fourth overtones of O-H and C-H stretching modes [8]. These wavelengths are at the vicinity of informative wavelengths of the third overtone of OH stretching (780nm) [28]; the third combination overtone of sugar OH stretching at 840nm (851nm) [29]; the strong absorption of water and oxygen (870nm) [30] that was used to predict the SSC of citrus [31]; the third overtone of CH stretching (910nm) [32]; and the strong water absorption band for the second and third overtone of the OH stretching (940nm) [33]. The OPT 101 has an on-chip amplifier to improve the signal-to-noise of the acquired signals by eliminating the noise pick up and leakage current error during data acquisition. The measured voltage of the photodiode is proportional to the detected diffuse reflectance intensity from the samples. The photodiode was allocated perpendicular to the sensor head where the sensor head was utilized to scan the pineapple. Figure 2(a) illustrates the framework of the in-house developed pre-dispersive NIR devices. The angle between the LEDs and the photodiode was set to 45°. Detail explanation of this pre-dispersive device was reported in [9].

Figure 2(b) illustrates the design of the in-house developed post-dispersive NIR device. It was developed using an AS7263 Near Infrared (NIR) Spectral Sensor (ams AG, Austria). Similar previous works that used the same sensor for different applications were [34], [35]. The on-board 5700k extremely bright white LED was used with maximum current supply of 100mA that gave the maximum light intensity to the sample and the spectral sensor can detect the intensity of reflected light at 680, 730, 760, 810, and 860nm. The voltage of the spectral sensor is proportional to the detected light intensity from a sample. A bifurcated fiber optic probe was used to guide the NIR energy (from the LED) toward the sample, and simultaneously to collect the diffuse reflected energy from the sample using the y-shaped concept in a single probe. The y-shaped concept is defined as the cable of light source and the cable of detection are mounted under a single sensor head. Thus, there is no angle between the light source and detector. Besides, the sensor head was directly contacted on the surface of a pineapple. Since there was no gap between the sample surface and the sensor head, the geometrical effect was minimized.
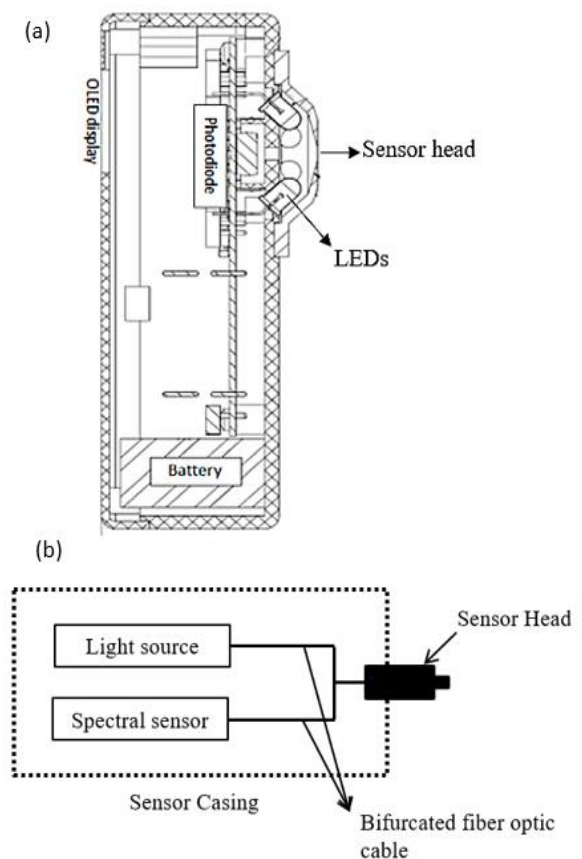

Fig. 2. The in-house developed pre- and post-dispersive design sensing devices: (a) the frame work of the pre-dispersive devices [9]; (b) The design of the post-dispersive devices.

2.2. Conventional Brix Measurement

Immediately after the non-destructively diffuse reflectance acquisition of a pineapple with both devices was completed, the Brix value of the pineapples was

measured. The flesh along with the peel that was at the selected "eye" was excised using a stainless-steel corer with a dimension of 50mm depth and 20mm diameter. The length of each specimen was limited to 20mm depth while the surplus flesh was removed. Then, the specimen without its 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 the juice. Lastly, few drops of the juice from the beaker were dropped on a digital refractometer (PAL-1, Atago, Tokyo, Japan) to obtain the Brix value. The entire process was repeated for another selected "eye" of the pineapple.

2.3. Data-preprocessing

During the diffuse reflectance collection, both devices were undergone white and dark calibration. To minimize the impact of surrounding unwanted light, the white and dark calibration values were taken under the same condition [36]. The values were then used to calculate the calibrated diffuse reflectance. The calibrated diffuse reflectance (I_c) was the ratio of the difference between the raw data (I_{led}) and the dark reference value (I_d), to the difference between the white reference (I_w) and I_d as that shown in Eq. (1) [37].

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

The calibrated diffuse reflectance was then pre-processed with two different methods i.e. normalization and standard normal variate (SNV). Normalization was carried out using Eq. (2). The normalized diffuse reflectance (I_N) is the ratio of the calibrated diffuse reflectance of each wavelength (I_c) to the total calibrated diffuse reflectance of the all the five wavelengths (I_{total}).

$$I_N = \frac{I_c}{I_{total}} \quad (2)$$

SNV is a popular pre-processing method that used to reduce the scattering effects and to correct the baseline shifts [38]. The SNV was performed using Eq. (3). The diffuse reflectance with SNV (I_{SNV}) was calculated by subtracting the calibrated diffuse reflectance data (I_c) of each wavelength by the average of the five calibrated diffuse reflectance ($\bar{X}$) and then divided by the standard deviation (σ) of those five wavelengths.

$$I_{SNV} = \frac{I_c - \bar{x}}{\sigma} \quad (3)$$

Using external validation data from different days are important in evaluating the robustness of NIR models. Nevertheless, for this work, we focused on the comparison between the pre- and post-dispersive NIR devices instead of the robustness of the models for future prediction. Thus, holdout validation was used, in which, the acquired data were holdout using Kennard-Stone (KS) algorithm as the testing data set (Table 1), while the

remainders were used to build the calibration model. The dataset was split into training (80%) with 576 data and testing (20%) with 144 data using KS algorithm. Similar training:testing ratio of 80:20 was used in the previous works e.g. 80:20 [39]–[41] and 77:23 [42]. Nevertheless, the ratio of training and testing data is not expected to affect this comparison study as the same training and prediction samples were used.

KS algorithm gains wide acceptance in splitting the original data sets into two different subsets [43],[44]. The algorithm is started by selecting two samples in which the Euclidean distance between them is the largest. The pair of samples was removed from the original dataset and put into the training set. Similar procedure is repeated until the number of samples required for training set is achieved. The Euclidean distance could be calculated using Eq. (4).

$$ED_x(p, q) = \sqrt{\sum_{j=1}^{N}[x_p(j) - x_q(j)]^2} \quad p, q \in [1, M] \quad (4)$$

where N is the number variables in x, and M is the number of samples. $x_p(j)$ and $x_q(j)$ are the j-th variable for p-th and q-th samples, respectively. The unselected samples were assigned as testing dataset. The advantage of KS algorithm is that the training set completely covers the measured region of the original set, and the testing samples are located inside the measured region.

2.4. Model Development

Three layers (i.e. one input, one hidden, and one output layers) artificial neural network (ANN) with Levenberg-Marquardt training algorithm was established via software MATLAB R2019a. The inputs of ANN were the acquired NIR diffuse reflectance signals and the output of ANN was the predicted soluble solids content (SSC). The architecture of the ANN is illustrated in Fig. 3.
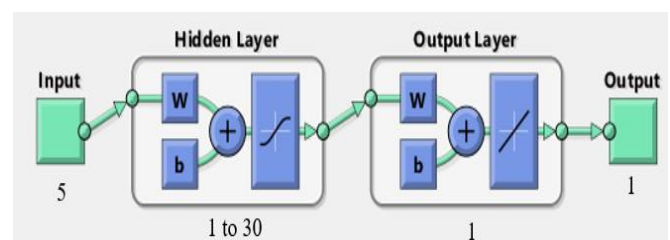

Fig. 3. The neural diagram of neural network: the inputs were the five NIR diffuse reflectance signals; the hidden neurons were varied from 1 to 30; and the output was the predicted SSC.

First, the five inputs (i.e. raw diffuse reflectance) and respective target (i.e. measured soluble solids content (SSC)) were mapped in the range between -1 and +1. Second, the training process was repeated 100 times with different random initial weights and biases, and different hidden neurons (varied from one to 30) to determine the optimal calibrated model. After that, the calibrated model was validated with the testing set. This training process

was repeated using different inputs of calibrated diffuse reflectance, calibrated diffuse reflectance with normalization, and calibrated diffuse reflectance with SNV. Since two NIR devices were involved, a total of eight ANN models were developed for this comparison study.

Table 1.   Descriptive statistics of the measured SSC of pineapples using a digital refractometer.

Dataset	Number of samples	Min (°Brix)	Max (°Brix)	Mean (°Brix)	Standard Deviation (°Brix)
Training	576	7.6	18	13.0	1.97
Testing	144	9.4	18	13.1	1.91
Total	**720**	**7.6**	**18.0**	**13.0**	**1.96**

Table 2.   The training and prediction performance of the ANN using the pre- and post-dispersive NIR devices.

Device	Pre-processing	HN	Training			Prediction				
			Bias	RMSEC (°Brix)	R_C	RMSEP (°Brix)	R_p	SEP	RPD	RER
Pre-dispersive	Raw	30	0.0000	1.4566	0.6745	1.6185	0.5251	1.6185	1.2172	6.4257
	CDR	**28**	**-0.0067**	**1.1305**	**0.8195**	**1.3918**	**0.6853**	**1.3918**	**1.4154**	**7.4724**
	CDR+N	23	-0.0029	1.3500	0.7290	1.5910	0.5648	1.5910	1.2382	6.5368
	CDR+SNV	27	-0.0094	1.4410	0.6828	1.7390	0.4398	1.7390	1.1329	5.9805
Post-dispersive	Raw	30	0.0027	1.1611	0.8084	1.5688	0.6294	1.5688	1.2557	6.6293
	CDR	4	0.000	1.3660	0.7216	1.3940	0.6882	1.3940	1.4132	7.4605
	CDR+N	4	0.000	1.4020	0.7033	1.4310	0.6739	1.4310	1.3767	7.2676
	CDR+SNV	**13**	**0.0042**	**1.2070**	**0.7911**	**1.3130**	**0.7408**	**1.3130**	**1.5004**	**7.9208**

Raw: Raw diffuse reflectance without pre-processing; CDR: Calibrated diffuse reflectance; CDR+N: Calibrated diffuse reflectance with Normalization; CDR+SNV: Calibrated diffuse reflectance with standard normal variate; HN: Hidden neuron.

2.5. Performance evaluation

In this study, the root mean squared error of calibration/prediction (RMSEC/RMSEP), correlation coefficient of calibration/prediction (R_c/R_p), bias, Standard Error of Prediction (SEP), Ratio of standard error of Performance to standard Deviation (RPD), and Range Error Ratio (RER) were calculated to evaluate the performance of the developed models.

The best calibration model should achieve the lowest RMSEP and the highest R_p [45]. The bias is defined as the averaged difference between the predicted and the real values of a calibration model. RPD is the ratio of the square root of the difference between the squared RMSEP and the squared bias to the standard deviation of the testing data. The prediction accuracy of a model is proportionally inclined with the RPD value [46]. Next, RER was the ratio of the range of prediction to the SEP. The RER value reflects the quality of a calibration model. For an RER > 4, the calibration model is acceptable for sample screening; for an RER > 10, the calibration model is acceptable for quality control; and for an RER > 15, the calibration model is good for quantification [47].

3. Results and Discussion

3.1. Descriptive Statistics of the SSC

Table 1 summarizes the descriptive statistics of 720 measured SSC from the selected "eyes" along the middle part of pineapples using the digital refractometer. The minimum and maximum of the measured SSC values were 7.6 and 18.0 °Brix, respectively. The measured SSC values were normally distributed with a mean value of 13.0 °Brix and a standard deviation of 1.96 °Brix. Since KS algorithm was used for the holdout validation, the standard deviation of the testing data was lower than that of the training data; while the mean values of both training and testing data sets were similar, i.e. 13.1 and 13.0, respectively.

3.2. Prediction Performance of ANN

Figures 4 (a) and 4 (b) illustrate the NIR raw spectra that acquired using the pre-dispersive and the post-dispersive NIR devices, respectively. Obvious baseline shifts that were observed should be due to the uneven surface of the pineapples. As NIR spectrum is highly overlapped and correlated, suitable pre-processing and machine learning are needed to remove unwanted signals and to establish a useful predictive model.
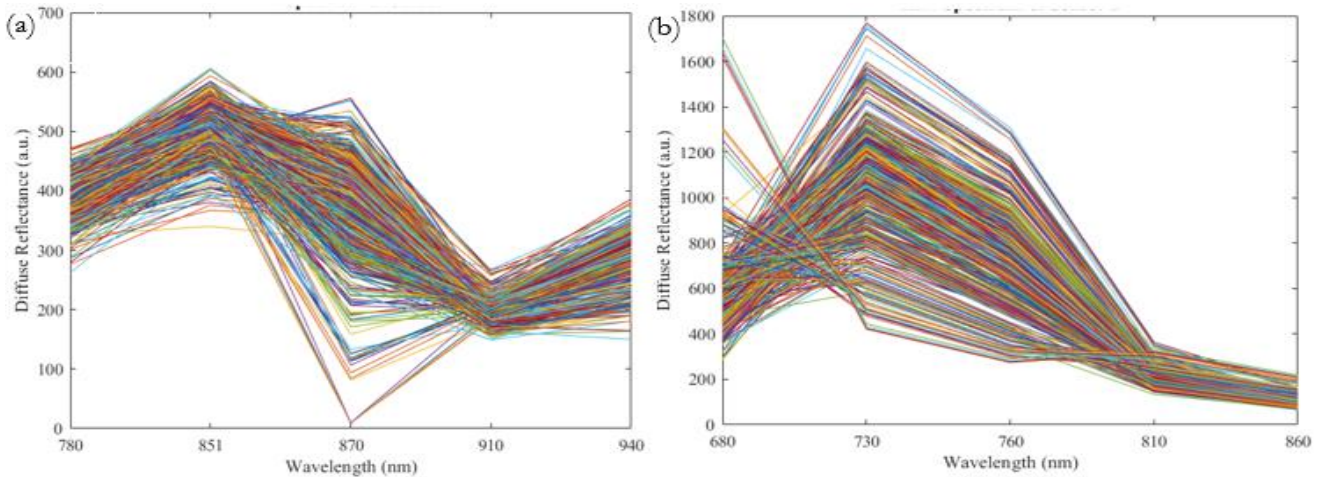
Fig. 4. The raw spectrum of devices: (a) pre-dispersive device, and (b) post-dispersive device.
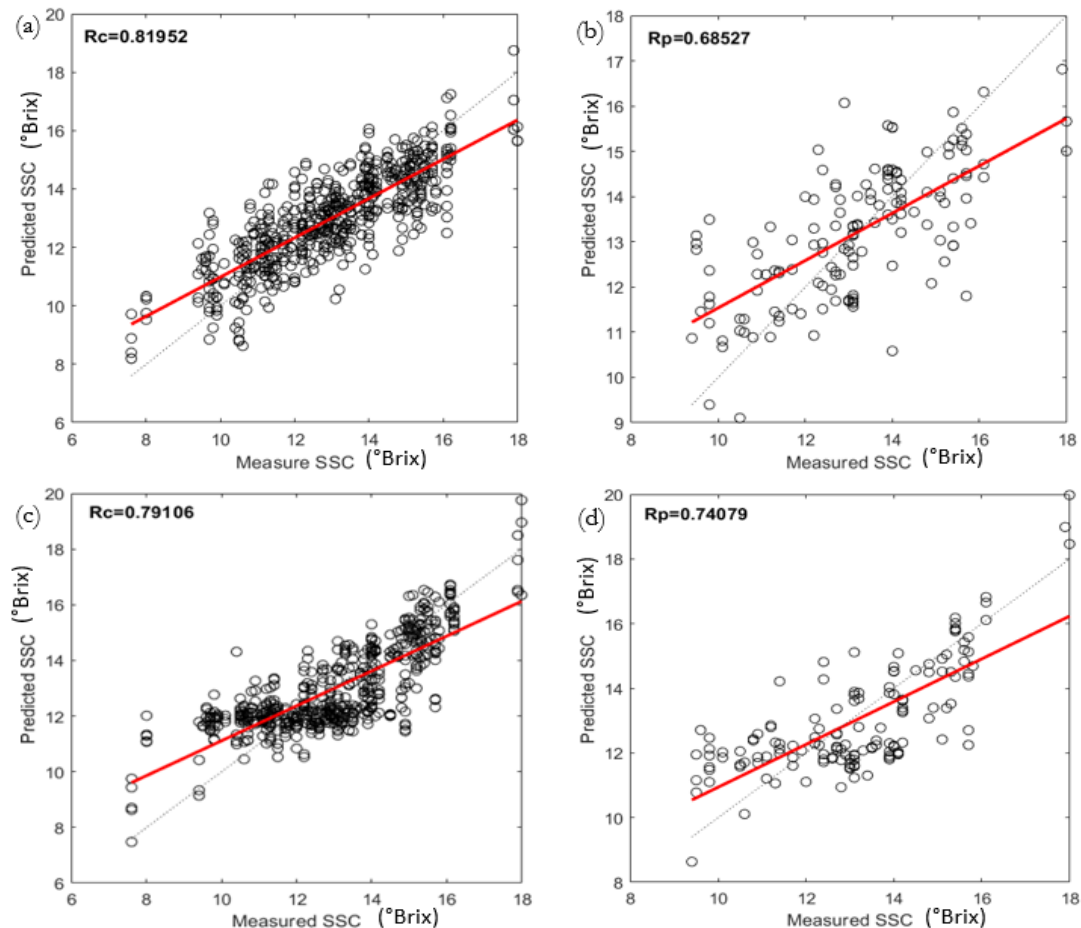

Fig. 5. The correlation plots of the optimal ANN model: (a) the ANN$_{CDR}$ of the pre-dispersive device for training, (b) the ANN$_{CDR}$ of the pre-dispersive device for testing, (c) the ANN$_{CDR+SNV}$ of the post-dispersive device for training, and (d) the ANN$_{CDR+SNV}$ of the post-dispersive device for testing.

Table 2 tabulates the training and testing performance of all calibration models with the diffuse reflectance collected by the two NIR devices i.e. pre- and post-dispersive in terms of RMSEC, RMSEP, Rc, Rp, SEP, RER, and RPD. The eight calibrated models (where each device consists of four calibration models) were built using the four different datasets i.e. raw diffuse reflectance (ANN$_{raw}$), calibrated diffuse reflectance (ANN$_{CDR}$), calibrated diffuse reflectance with normalization

(ANN$_{CDR+N}$), and calibrated diffuse reflectance with SNV (ANN$_{CDR+SNV}$), respectively.

The ANN that used raw spectral data (ANN$_{raw}$) for the pre-dispersive device revealed the lowest Rc of 0.6745 and the highest RMSEC of 1.4566 among the four calibrated models. Relative to the post-dispersive device, the R$_C$ of ANN$_{raw}$ gave the optimal performance among the other models with Rc of 0.8084 and the lowest RMSEP of 1.1611 °Brix. However, both models did not achieve the best prediction performance. Nevertheless, ANN with 30 hidden neurons implies that the complexity of spectra was high. The use of pre-processing not only reduced the number of hidden neuron that was needed to achieve the optimal performance, but also potentially gave better prediction performance by removing unwanted signals that can avoid overfitting issues.

For the pre-dispersive device, ANN$_{CDR}$, ANN$_{CDR+N}$, and ANN$_{CDR+SNV}$ achieved RMSEC and Rc of 1.1305 °Brix and 0.8195; 1.35 °Brix and 0.7290; and 1.441 °Brix and 0.6828, respectively. For the prediction accuracy evaluation, ANN$_{CDR}$, ANN$_{CDR+N}$, and ANN$_{CDR+SNV}$ achieved RMSEP and Rp of 1.3918 °Brix and 0.6853; 1.591 °Brix and 0.5648; and 1.739 °Brix and 0.4398, respectively. Notably, ANN$_{CDR}$ was the best model with the lowest RMSE and the highest R value in both training and testing evaluation. This shows that unstable pre-processing method might remove the interested information and consequently degraded the prediction performance. This is aligned with the previous work in particleboard types classification, in which, the spectra without pre-processing and the normalized spectra gave the same values with 99.1% whereas the spectra that pre-processed by SNV was 0.9% lower than both of them [48].

On the other hand, the prediction performance of ANN$_{CDR+N}$ was better than ANN$_{CDR+SNV}$ in pre-dispersive deign. The result is in-line with the previous SSC prediction of sugar beet juices study, in which, the SNV gave the R$_p{}^2$ of 0.983; while the normalization gave the R$_p{}^2$ with 0.984 [49]. Besides, the result from rapid prediction and visualization of moisture content in single cucumber also showed that the prediction performance of normalization (Rp = 0.918) was better than the performance of SNV (Rp = 0.911).

In previous study about the SSC prediction of pineapple, the RMSEP achieved with range between 0.82 to 1.01 °Brix and the R$_p$ ranged 0.68 to 0.74 [50]. Conversely, the RMSEP values in this study was relatively higher with range 1.39 to 1.74 °Brix. Besides, the R$_p$ of the pre-dispersive models ranged from 0.4398 to 0.6853 was much lower than the previous study. This degraded RMSEP and R$_p$ might be caused by the lack of wavelength range covered, and subsequently the small portion of meaningful information was absent. Related study about the selection of specific wavelengths and the loss of information in NIR data had been reviewed using a simplified LED design with four NIR wavelengths in rapid ripeness evaluation of white grape [29].

Table 2 shows that the RPD value of the four pre-dispersive ANN models ranged from 1.13 to 1.41 where the highest RPD value i.e. 1.41 was achieved by the ANN$_{CDR}$. On the other hand, the range of RPD value for the four post-dispersive ANN models varied from 1.25 to 1.50, and the ANN$_{CDR+SNV}$ model gave the highest values with 1.50. Even though the RPD values for both designs were relatively low, the post-dispersive design shows a better prediction accuracy with higher RPD value. Nevertheless, the RER values of both pre- and post-dispersive model suggest that both NIR devices were acceptable for sample screening where the range of value achieved 5.98 to 7.47 and 6.63 to 7.92, respectively. Magale Karine Diel RAMBO et al. reported near infrared spectroscopy (NIRS) that was able to achieve RER > 5 could be useful to screen the quality of solid intact sample of coconut and coffee husks [51].

Interestingly, ANN$_{CDR}$ achieved similar RPD and RER for both pre- and post-dispersive NIR device. The involvement of SNV degraded the accuracy of pre-dispersive NIR device; but that improved the accuracy of post-dispersive NIR device. Consequently, the best models with the highest RER values (i.e. 7.47 and 7.92) were ANN$_{CDR}$ and ANN$_{CDR+SNV}$ in pre- and post-dispersive NIR devices, respectively. This suggests that different acquisition setups require different pre-processing methods to reveal its potential.

3.3. Post-dispersive design

For the post-dispersive device, the RMSEC and R$_c$ of three different models i.e. ANN$_{CDR}$, ANN$_{CDR+N}$, and ANN$_{CDR+SNV}$ were 1.366 °Brix, 0.7216; 1.42 °Brix, 0.7033; and 1.207 °Brix, 0.7911, respectively; while the RMSEP and Rp were 1.394 °Brix; 0.6882, 1.431 °Brix; 0.6739, and 1.313 °Brix; 0.7408 respectively. Among these models, the ANN$_{CDR+SNV}$ had revealed the optimal training and prediction performance with lowest RMSE and highest R values.

ANN$_{CDR+SNV}$ outperformed the ANN$_{CDR}$ and ANN$_{CDR+N}$ was in-lined with the previous study in investigating the parameters that affected the acquired diffuse reflectance in non-destructively SSC prediction of pineapple [11]. Besides, the study in rapid estimation of leaf nitrogen content in apple-trees also showed that the ANN$_{CDR+SNV}$ gave the best prediction performance compared with others pre-processing method including normalization [52]. The degraded prediction performance of ANN$_{CDR+N}$ might be due to the similar reason in the pre-dispersive models where the pre-processing method removed the meaningful information from the spectrum data.

3.4. Comparison between Pre- and Post-dispersive

The pre-dispersive models achieved RMSEP ranged between 1.392 to 1.739 °Brix, whereas, the RMSEP of the post-dispersive models were relatively lower with a range from 1.313 to 1.43 °Brix. The lower range of RMSEP

might be caused by the stability of a bifurcated fiber optic design in minimizing the geometrical effects on the uneven surface of the pineapple. Consequently, the R_p values of the post-dispersive models that ranged from 0.6882 to 0.7408 was higher than that of the pre-dispersive models (that ranged from 0.4398 to 0.6853). This is worth to highlight that the post-dispersive model achieved similar R_p that reported from the previous study in SSC prediction of pineapple, in which, the range of Rp is 0.68 to 0.74 [50].

Next, the ANN of the post-dispersive device required fewer hidden neurons to achieve its optimal performance compare to that of the pre-dispersive device. This suggests that the spectra that collected on post-dispersive were more consistent than the spectra that collected on pre-dispersive device. This might be due to the uneven surface of a pineapple that affects the travel path of the light. This unwanted effect was minimized using the bifurcated fiber optic design that eliminated the geometrical effects on the uneven surface of pineapples. Consequently, the correlation coefficient of the optimal post-dispersive model outperformed the optimal pre-dispersive model with an improvement of an approximate 8.1% (i.e. R_p from 0.6853 to 0.7408, in which, ((0.7408 - 0.6853)/0.6853)*100% = 8.1%), and the RMSEP was reduced by 5.7% (i.e. from 1.3918 to 1.3130°Brix, in which, ((1.3918 - 1.3130)/1.3918)*100% = 5.7%).

3.5. Comparison between the optimal models

Figure 5 illustrates the correlation plot of the optimal models that built using the pre- and post-dispersive devices, in which, Fig. 5(a) and Fig. 5(b) are the training and testing of ANN_{CDR} for the pre-dispersive device; Fig. 5(c) and Fig. 5(d) are the training and testing of $ANN_{CDR+SNV}$ for the post-dispersive device.

Figures 5(a) and 5(c) illustrate that the training performance of the optimal model of both devices is satisfactory whereas the performance of the post-dispersive device could be better if the training data around the minimum and maximum values gave a lower calibrated error. For instance, the training data at the 8 and 18 °Brix were not well calibrated that consequently degraded the training performance. Conversely to the training performance, the prediction performance of the pre-dispersive optimal model was worse than the post-dispersive optimal model. This indicates that the pre-dispersive model was too optimistic with higher training performance; while the post-dispersive model was more robust to have a lower RMSEP when it was tested using the testing data. This implies that the pre-dispersive model that required more hidden neurons to achieve its optimal performance reduced the robustness of the model.

4. Conclusion

In this study, the pre- and post-dispersive NIR devices were utilized in diffuse reflectance collection. The effects of these designs on the prediction accuracy were evaluated using artificial neural network (ANN). Eight calibrated models were built and compared using four different spectral data i.e. raw diffuse reflectance (ANN_{raw}), calibrated diffuse reflectance (ANN_{CDR}), calibrated diffuse reflectance with normalization (ANN_{CDR+N}), and calibrated diffuse reflectance with SNV ($ANN_{CDR+SNV}$), respectively, for both devices.

The optimal model of the pre-dispersive device achieved RMSEP of 1.3918 °Brix and R_p of 0.6853; while that of the post-dispersive device obtained a lower RMSEP of 1.313°Brix and a higher R_p of 0.7408. Furthermore, the number of hidden neurons that was used for the ANN of the post-dispersive was fewer than that was used for the ANN of the pre-dispersive device. This suggests that the prediction performance of the pre-dispersive device might be affected by the geometrical effects on the uneven surface of pineapples.

On the other hand, the mentioned effect was minimized when the post-dispersive device coupled with a bifurcated fiber optic probe was used. Notably, with the post-dispersive design, the correlation coefficient of the optimal ANN was increased with approximately 8.1% i.e. R_p from 0.6853 to 0.7408, and the RMSEP of that was reduced by 5.7% from 1.3918 to 1.313°Brix. The optimal post-dispersive model achieved RER of 7.92 and RPD of 1.50 that indicates that the model has acceptable prediction accuracy for sample screening.

Next, the findings show that SNV degraded the prediction accuracy of the pre-dispersive NIR device but it improved the prediction accuracy of the post-dispersive NIR devices. Thus, a better Brix prediction performance can be achieved with a proper NIR measurement setup with a suitable pre-processing method. A better understanding about the geometrical effects on different fruits is important to produce a better sensing device design, subsequently, to increase the prediction accuracy. Future works may study the effects of different calibration algorithms and pre-processing methods to address the potential unwanted geometrical effects on different fruits that have different types of surface characteristics.

Acknowledgement

References

[1] Y. Liu *et al.*, "Quality evaluation of fried soybean oil base on near infrared spectroscopy," *J. Food Process Eng.*, vol. 41, no. 7, p. e12887, 2018.

[2] 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, pp. 170–177, 2018.

[3] Y. Xia, W. Huang, S. Fan, J. Li, and L. Chen, "Effect of fruit moving speed on online prediction of soluble solids content of apple using Vis/NIR diffuse transmission," *J. Food Process Eng.*, vol. 41, no. 8, pp. 1–10, 2018.

[4] B. Li *et al.*, "Pears characteristics (soluble solids content and firmness prediction, varieties) testing methods based on visible-near infrared hyperspectral imaging," *Optik (Stuttg).*, vol. 127, no. 5, pp. 2624–2630, 2016.

[5] 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.

[6] 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.

[7] K. Kamwilaisak, P. Jutakridsada, and W. Iamamornphanth, "Estimation of Sugar Content in Sugarcane and Khon Kaen 3 (KK 3) by Near Infrared Spectroscopy," *Eng. J.*, vol. 25, no. 3, 2021.

[8] J. Posom *et al.*, "Two different portables visible-near infrared and shortwave infrared region for on-tree measurement of soluble solid content of marian plum fruit," *Eng. J.*, vol. 24, no. 5, pp. 227–236, 2020.

[9] K. S. Chia, M. N. H. Jam, Z. Gan, and N. Ismail, "Pre-dispersive near-infrared light sensing in non-destructively classifying the brix of intact pineapples," *J. Food Sci. Technol.*, vol. 57, no. 12, pp. 4533–4540, 2020.

[10] J. Guthrie and K. Walsh, "Influence of environmental and instrumental variables on the non-invasive prediction of Brix in pineapple using near infrared spectroscopy," *Aust. J. Exp. Agric.*, vol. 40, no. 2, pp. 681–696, 2000.

[11] F. W. Hong and K. S. Chia, "Investigation of parameters that affect the acquired near infrared diffuse reflected signals in non-destructive soluble solids content prediction," *Eng. J.*, vol. 24, no. 6, pp. 79–90, 2020.

[12] M. Zareef *et al.*, "An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis," *Food Eng. Rev.*, vol. 12, no. 2, pp. 173–190, 2020.

[13] N. T. Anderson, K. B. Walsh, P. P. Subedi, and C. H. Hayes, "Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content," *Postharvest Biol. Technol.*, vol. 168, p. 111202, 2020.

[14] D. X. Guo, Y. Z. Shang, R. Peng, S. S. Yong, and X. A. Wang, "Noninvasive blood glucose measurement based on NIR spectrums and double ANN analysis," *J. Biosci. Med.*, vol. 3, no. 6, pp. 42–48, 2015.

[15] S. D. Afandi, Y. Herdiyeni, L. B. Prasetyo, W. Hasbi, K. Arai, and H. Okumura, "Nitrogen content estimation of rice crop based on Near Infrared (NIR) Reflectance using artificial neural network (ANN)," *Procedia Environ. Sci.*, vol. 33, pp. 63–69, 2016.

[16] V. R. Sharabiani, A. S. Nazarloo, and E. Taghinezhad, "Prediction of protein content of winter wheat by canopy of near infrared spectroscopy (NIRS), using partial least squares regression (PLSR) and artificial neural network (ANN) models," *Yuz. Yil Univ. J. Agric. Sci.*, vol. 29, no. 1, pp. 43–51, 2019.

[17] A. Puttipipatkajorn and A. Puttipipatkajorn, "Development of calibration models for rapid determination of moisture content in rubber sheets using portable near-infrared spectrometers," *J. Innov. Opt. Health Sci.*, vol. 13, no. 2, p. 2050009, 2020.

[18] X. Zhang, L. Chen, Y. Sun, Y. Bai, B. Huang, and K. Chen, "Determination of zinc oxide content of mineral medicine calamine using near-infrared spectroscopy based on MIV and BP-ANN algorithm," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 193, pp. 133–140, 2018.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] I. Bhati and M. Bhattacharya, "An IOT-based system for classification and identification of plastic waste using near infrared spectroscopy," *Proc. 2nd Int. Conf. Commun. Devices Comput. Lect. Notes Electr. Eng.*, vol. 602, pp. 697–703, 2020.

[35] D. Durmus and W. Davis, "Multi-channel low-cost light spectrum measurement using a multilayer perceptron," *Energy Build.*, vol. 199, pp. 579–587, 2019.

[36] 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.

[37] 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.

[38] 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.

[39] U. Inyaem, "Construction model using machine learning techniques for the prediction of rice produce for farmers," in *2018 IEEE 3rd Int. Conf. Image, Vis. Comput.*, 2018, pp. 870–874.

[40] Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, "A Review of the artificial neural network models for water quality prediction," *Appl. Sci.*, vol. 10, no. 17, 2020.

[41] K. H. Anita Rácz, Dávid Bajusz, "Effect of dataset size and train / test split ratios in QSAR/QSPR multiclass classification," *Molecules*, vol. 26, no. 4, pp. 1–16, 2021.

[42] 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.

[43] 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, pp. 39–45, 2018.

[44] H. Tian et al., "Weighted SPXY method for calibration set selection for composition analysis based on near-infrared spectroscopy," *Infrared Phys. Technol.*, vol. 95, pp. 88–92, 2018.

[45] 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.

[46] N. Cristina, T. Mariani, G. Henrique, D. A. Teixeira, K. Michell, and G. De Lima, "NIRS and i SPA – PLS for predicting total anthocyanin content in jaboticaba fruit," *Food Chem.*, vol. 174, pp. 643–648, 2015.

[47] "NIR Model Development At Celignis." Celignis.com. [Online]. Available: https://www.celignis.com/NIR.php (accessed: Aug. 13, 2021).

[48] A. Sandak, J. Sandak, D. Janiszewska, S. Hiziroglu, M. Petrillo, and P. Grossi, "Prototype of the near-infrared spectroscopy expert system for particleboard identification," *J. Spectrosc.*, vol. 2018, no. 1, pp. 1–11, 2018.

[49] M. E. Bahrami, M. Honarvar, K. Ansari, and B. Jamshidi, "Measurement of quality parameters of sugar beet juices using near-infrared spectroscopy and chemometrics," *J. Food Eng.*, vol. 271, p. 109775, 2020.

[50] 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.

[51] M. Karine et al., "Prediction of quality parameters of food residues using NIR spectroscopy and PLS models based on proximate analysis," *Food Sci. Technol.*, vol. 40, no. 2, pp. 444–450, 2020.

[52] S. Chen *et al.*, "Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods," *Infrared Phys. Technol.*, vol. 111, p. 103542, 2020.

Fan Wei Hong was born in Batu Pahat, Johor, Malaysia in 1994. He obtained his MEng (electrical engineering) and BEng (mechatronic and robotic engineering) from Universiti Tun Hussein Onn Malaysia (UTHM) at Batu Pahat, Malaysia in Sept.2021 and June. 2018, respectively. 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 has professionally registered with Engineering Council as Chartered Engineers (CEng) through The Institution of Engineering and Technology (IET) since Nov. 2020.

He was an assistant professor at Southern University College, Skudai, Malaysia from March 2014 to August 2014. He has served as a senior lecturer in the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia since September 2014. He has been the principal investigator of more than 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.

Xien Yin Yap was born in Lahad Datu, Sabah, 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. During his degree, he was a winner of Imagine Cup Asia Pacific Regional Finals and the project was related to machine learning and near infrared spectroscopic analysis. From 2018 – 2021, he was pursuing his MEng in electrical engineering in Faculty of Electrical and Electronic Engineering, UTHM. His research interests are machine learning, artificial intelligence and near infrared spectroscopic analysis.