AgroLife Scientific Journal - Volume 10, Number 1, 2021
ISSN 2285-5718; ISSN CD-ROM 2285-5726; ISSN ONLINE 2286-0126; ISSN-L 2285-5718

GEOGRAPHICAL IDENTIFICATION OF DURIAN (CV. MONTHONG) FROM PRACHUAP KIRI KHAN AND CHANTHABURI PROVINCE BY USING NEAR INFRARED SPECTROSCOPY

Kingdow CHANACHOT[1,2], Wanphut SAECHUA[1,2], Panmanas SIRISOMBOON[1,2]

[1]Agricultural Engineering Department, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, 1 Chalong Krung 1 Alley, Ladkrabang, Bangkok, Thailand
[2]NIR Spectroscopy Research Center for Agricultural Product and Food, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, 1 Chalong Krung 1 Alley, Ladkrabang, Bangkok, Thailand

Corresponding author emails: kingdow_c@hotmail.com, wanphut.sa@kmitl.ac.th

Abstract

Geographic origin of durian (Durio zibethinus Murray) can influence consumers purchasing decisions. A guaranteed geographical origin is extremely useful to maintain the credibility and popularity of this fruit. The objective of this work is to confirm geographical identification (GI) of durian Monthong from Prachuap Kiri Khan Province from south of Thailand which achieved the GI lable from the department of intellectual property by near infrared spectroscopy (NIRs). Total of 60 durian Monthong samples were acquired in the wavenumber range of 12,500-4,000 cm^{-1}by FT-NIR spectrometer. The sample presentation for scanning were 3 forms including intact peel (with thorns), cut-thorns peel and stem of the fruit. Soft independent modeling of class analogy (SIMCA), partial least squares discriminant analysis (PLS-DA) and support vector machine classification (SVMC) were used to establish identification model using the full spectra. Overall classification accuracy of SVMC was the highest in training and validation set of stems which was 100% and 100%, respectively. This study indicated that NIRs might be a method to guaranteed geographic origin of durian Monthong from Prachuap Kiri Khan. However, more robust model should be developed by using more samples.

***Key words**: durian Monthong, geographic origin, geographical identification.*

INTRODUCTION

Geographic origin of durian (*Durio zibethinus* Murray) can influence consumers purchasing decisions. A guaranteed geographical origin is extremely useful to maintain the credibility and popularity of this fruit. Different varieties or geographical origins of durian fruits offers a different flavour and texture. For the geographical origin, the observed differences were due to seasons, climates, soil minerals, irrigation, fertilizer, spraying, and planting method. These factors were controlled by the farmers guided by durian fruit experts. The price of durian depends on the season and its geographical origin. Both these two variables causes variation in texture and flavor, which is cannot be visually identified. The geographical indication (GI labeling) of durian from difference province in Thailand, has its own trademark and patent law from Department of Intellectual Property (DIP), which is under the supervision of Ministry of Commerce. The label could bestated as a local brand that indicates the quality and source of the product. As well as durian plants in Malaysia, Chanachot et al. (2020) indicated the geographic origin of durian in Penang, Malaysia. Therefore, the geographic origin should be identified by analytical methods, otherwise the false identification of the fruit's geographical origin could have a negative impact on the image of that country's export products (Chanachot et al., 2020). This way the GI label, it might have subrogation.

Currently the discrimination method of durian geographical origin is limited to the detection of a few active components, using analytical methods. DNA is an abbreviation for deoxyribonucleic acid, any living thing that makes up strands of DNA which the once method that can identify durian cultivars (Sales, 2015). This method is time-consuming, uses chemicals and requires special human skill.

The one method which could be used to identify the geographic origin of durian Monthong from Prachuap Kiri Khan Province is NIR spectroscopy. The resultant spectrum contains information specific to the molecular vibrational aspects of the sample, its physical properties, and its unique interaction with the measuring instrument (Workman & Weyer, 2012). Relating the spectra to the chemical structure of the measured samples is referred to as spectra-structure correlation. The exclusive use of chemometrics alone provides a weak basis for analytical science (Workman & Weyer, 2012). Nowadays, Traditional Chinese Medicine (TCM) is using modern NIR spectroscopy in combination with chemometric methods to offer reliable species authentication and accurate geographical origin discrimination of herbal medicine. It is expected NIR spectroscopy in combination with chemometric methods to be further employed in the authentication and quality control of herbal medicines (Wang & Yu, 2015). For several herbal medicines: *Rhizoma corydalis* (Lai & Kokot, 2011), *Ganoderma lucidum* (Chen et al., 2008), *Angelicae gigantis radix* (Woo et al., 2015), *Glycyrrhizia uralensis* Fisch (Wang et al., 2007), *Scrophulariae radix* (Lee et al., 2011), *Peucedanum peucedanum* (Zhu et al., 2011) and *Paeoniae radix* (Luo et al., 2008), which were classified by discriminant analysis (DA), a similar model analysis of class analysis (SIMCA) (Chen et al., 2008) and partial discriminant analysis of the smallest squares (PLS-DA) was used. In studies of herbal medicines, the combination of classical analytical methods with NIR can provide the necessary accuracy, with a correct discrimination from acceptable up to 100%. Present, both agricultural product and food are characterized using NIRs. There are studies which are using NIR for to distinguish between peach varieties (Guo et al., 2016), to identify the geographical origin of apples (Eisenstecken et al., 2019) and Goji berry (Tingting et al., 2016). Most recently, NIRs combined with a support vector machine classification (SVMC) were used for the classification of almonds (*Prunus dulcis* MILL.), which showed a classification accuracy of 80.2% (± 1.9 %) of the validation set for determining the origin of freeze - drying almonds (Arndt et al., 2020).

The objective of this work is to confirm the geographical identification (GI) of durian, cv. Monthong, from Prachuap Kiri Khan Province from south of Thailand, which achieved the GI label separated from the durian from Chanthaburi by using NIRs, which is a method with the advantages of fast analytical speed, easy operation, non-destructive measurement, and dependability.

MATERIALS AND METHODS

Durian sample collection

Durian samples of Prachuap Kiri Khan Province from South of Thailand and Chanthaburi Province from East of Thailand were randomly picked from 5-10-year-old durian trees in 3 local orchards in each province with different owners.

The conditions of picking up durian samples is determined by National Bureau of Agricultural Commodity and Food Standards (Bureau of Commercial Provincial Office of the Permanent Secretary Ministry of Commerce, 2020). Fruits were harvest 120 days after anthesis (DAA) and the fruit weight was between 2.5-3.5 kg. The plants were irrigated, fertilized and sprayed by their owners guided by durian fruit experts. Totally 60 durian samples were transported direct to the NIRs Research Centre of Agriculture Product and Food, Department of Agricultural Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand. Then, kept at approximately 25ºC room temperature overnight prior to further measurements. All measurements were undertaken at room temperature (25 ± 2ºC).

NIR spectra collection

NIR spectral data were acquired in the wavenumber range of 12500-4000 cm^{-1} in diffuse reflectance mode with scanning resolution of 16 cm^{-1} in absorbance mode and there were 32 scans for 1 average spectrum of the sample by the FT-NIR spectrometer (Multi-Purpose Analyzer, MPA, Bruker optics, Leipzig, Germany). The surfaces of whole durian samples were chosen at the biggest 3-4 of locules at the middle of locule of the whole sample were chosen for scanning (3 times scanning per position). Black sponge was used to cover between sample and light beam source from spectrometer to avoid leaking of light. The sample presentation for scanning were 3 forms including intact peel (with thorns), cut-thorns peel and stem of the fruit.

Chemometric methods

Durian samples were divided into training set and validation set in ratio 2/3 and 1/3, respectively. Multivariate analysis including principal component analysis (PCA), soft independent modeling of class analogy (SIMCA), partial least squares discriminant analysis (PLS-DA), and support vector machine classification (SVMC) were applied to extract relevant features from the spectral data and to remove the anomalous samples. Soft independent modeling of class analogy (SIMCA) calculated a PCA model for each class. The unknown sample will be identified and sorted by class by comparing the residual variance of the modelled class with the residual variance of the unknown samples and the optimal number of PCs must be selected (Luna et al., 2013). The best classification performance of SIMCA is obtained when the distance between data is the largest (Luna et al., 2013). Partial least squares discriminant analysis (PLS-DA) was also used. A PLS regression model relates the independent variables (here spectra) to an integer y that designates the class of the sample. For example, the number one (1) is used to indicate that the training sample belongs to the class of interest, and a zero (0) indicates that the sample belongs to a different class. Classification of an unknown sample is derived from the value predicted by the PLS model, $\hat{y}$ (Arndt et al., 2020). This value is a real number, not an integer, which should be ideally close to the values used to codify the class (here either 0 or 1). A cut-off value between 0 and 1 is established so that a sample is assigned to class 1 if the prediction is larger than the cut-off value or assigned to class 0 otherwise (Arndt et al., 2020). The simplest approach is to use an arbitrary cut-off value, such as 0.5. SVMC is a classification method based on statistical learning wherein a function that describes a hyperplane for optimal separation of classes is determined (Camo Software AS, Oslo, Norway). Classification of an object is carried out strictly by assigning it to the class on either side of the separating plane (hyperplane). To deal with overlapping classes, one approach is to allow for some objects to be on the wrong side of the margin (Otto, 2016). A kernel function is used to map from the original space to the feature space, and can be of many forms, thus providing the ability to handle nonlinear classification cases (Camo Software AS, Oslo, Norway). Classification analysis was carried out using Unscrambler 9.8 software (Camo Software AS, Oslo, Norway). The % accuracy and % correct classification rate (% CCR) were calculated as follows: Accuracy = [number of (TP+ TN)/number of (TP+ TN+ FP + FN)] × 100%; TP is true positive, which is the event that a positive sample is classified as a positive example, TN is true negative, which is the event that a negative sample is classified as a negative example, FP is false positive, which is the event that a negative sample is classified as a positive example, FN is false negative, which is the event that a positive sample is classified as a negative example; and % CCR = [(correct classify samples/total of samples) × 100%]. Both % accuracy and % CCR were calculated to better mode in training set and validation set.

RESULTS AND DISCUSSIONS

Exploratory data analysis

Raw spectra features, the absorption of Prachuap Kiri Khan and Chanthaburi Province form of samples; intact peel (with thorns), cut-thorns peel and stem form. Most of high peak were dominated peak of water (at 10314, 8516, 6897, and 5230 cm^{-1}). A small peak at 7413 cm^{-1} (around 1360 nm) was due to the absorption band associated with the 2 x C−H str. + C−H def. of CH_3. The peak at 5585 cm^{-1} (around 1780 nm) was due to the absorption band associated with the C−H str. first overtone of cellulose and the peak at 4000 cm^{-1} (2500 nm) was absorption band of C−H str. + C−C str. of starch (Figure 1a) (Osborne, 2006). By visual comparison of average raw spectra of durian sample in all forms, no significantly spectral differences were found. Todiscriminate sample according to the geographical origin is very difficult just visual examination. Accordingly, NIR spectroscopy combined with pattern recognition technique for discriminate of origin is necessary (Duan et al., 2014). For spectral comparison, the Savitzy-Golay 2nd Derivative, 2 Polynomial order with 11 smoothing points were examined into all forms.

The corresponding NIR spectra as shown in Figure 1b. First small peak is 8709 cm^{-1} (around 1148 nm), was due to the absorption band associated with C−H str. second overtone of CH_3. At 7112 cm^{-1} (around 1410 nm) was due to the absorption band associated with C−H combination, CH methylene of lignin. At 5268 cm^{-1} (1900 nm) was due to the absorption band associated with O−H str. + 2 × CO str. of starch and the most significant difference is observed at 7151 cm^{-1} (around 1395 nm) and 8694 cm^{-1} around 1152 nm) (Figure 1c). At 1395 nm might owing to the band of lignin.

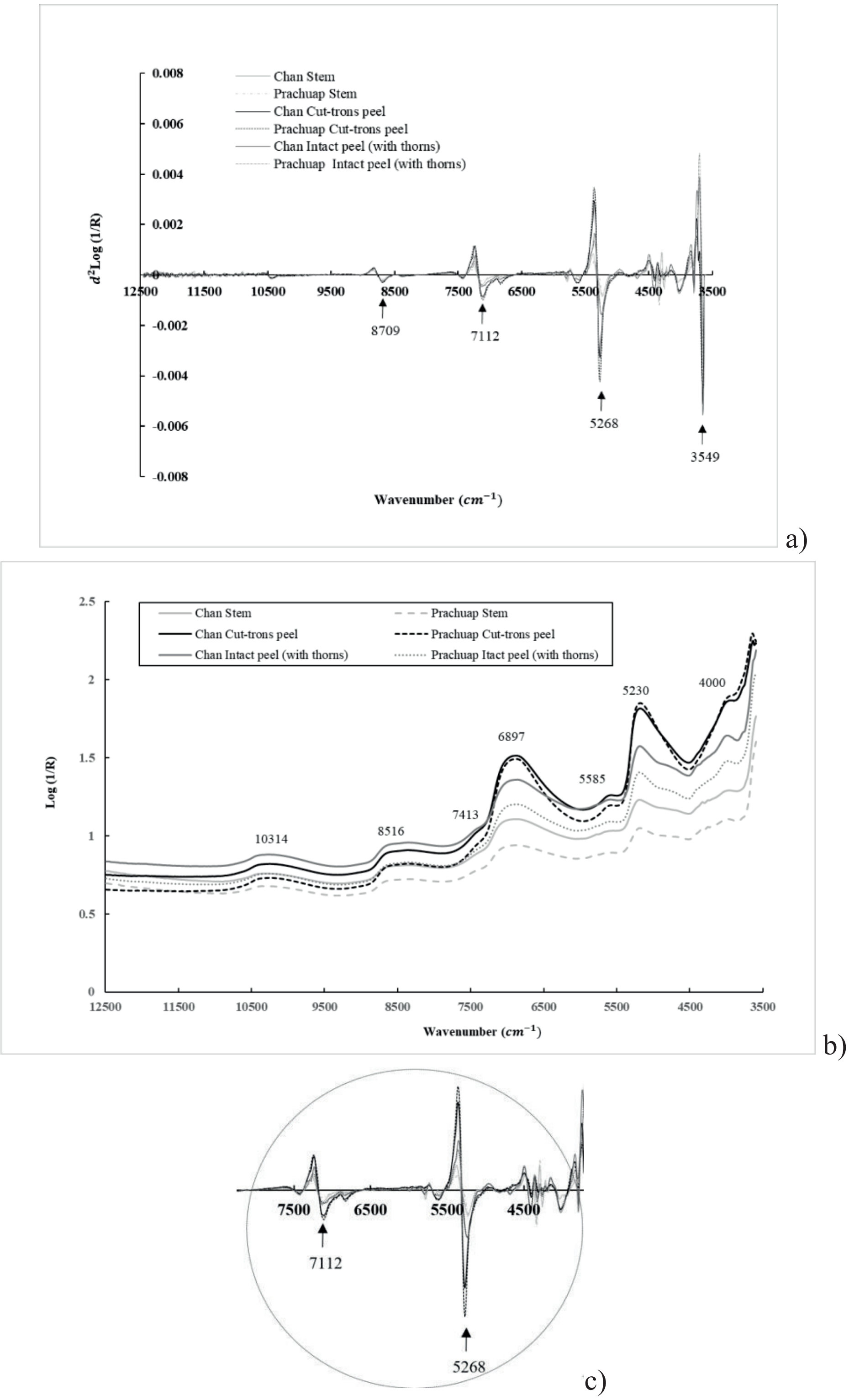

Figure 1. a) The average raw spectra of durian Monthog from Prachuap and Chanthaburi province in intact peel (with thorns), cut-thorns peel, and stem form; b) second derivative spectra of selected wavelength in the 9504 ÷ 3594 cm^{-1} range of stem form; c) expansion of Figure 1b)

In cut-thorns peel form, Prachuap Kiri Khan Province was higher than Chanthaburi Province but in intact peel (with thorns) and stem form, Chanthaburi was higher. In contrast, at peak 1900 nm might owing to the band of starch. Chanthaburi Province was higher than Prachuap Kiri Khan Province in cut-thorns peel form. For in cut-thorns peel and stem were nearly the same. These peaks can be assigned as the CH_2 and CH_3 band in chemical composition of stem (Osborne, 2006). These might cause in term of classification and its NIR bands can be successfully utilized in this study.

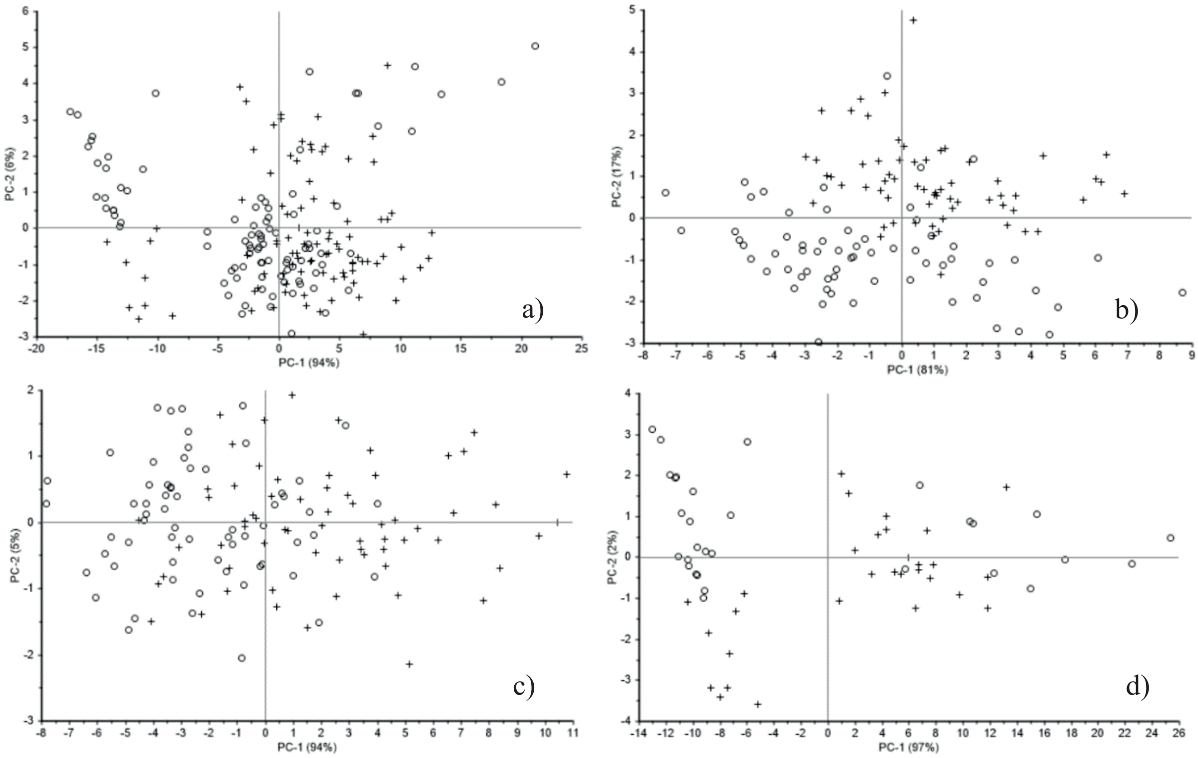

Figure 2. PCA result of all forms by raw spectra. a) Sample score (PC-1 vs. PC-2) of intact peel (with thorns) form; b) Sample score (PC1 vs. PC2) of cut-thorns peel form; c) Sample score (PC1 vs. PC2) of stem form; d) Sample score (PC1 vs. PC2) of combine form

Figure 2a shows PCA result of intact peel (with thorns) forms by raw spectra. PC-1 and PC-2 could explain total of variances to 94% and 6% respectively. Figure 2b shows PCA result of cut-thorns peel. PC-1 and PC-2 could explain total of variances to 81% and 17%, respectively. Figure 2c shows PCA result of stem. PC-1 and PC-2 could explain total of variances to 94% and 5%, respectively. And Figure 2d shows PCA result of combine. PC-1 and PC-2 could explain total of variances to 97% and 2%, respectively. The figure of PCA in all form of sample base on two groups according to their geographical origin are not clearly identified by visualization.

To relate the durian sample distribution observed in the score plot to spectral features, the loading plot of PC-1, PC-2 and PC-3 are represented in Figure 3. This was the X-loading of stem form because it outperformed higher % accuracy in overall classification. It was found the loading value on PC-2. In detail, variables are corresponding to the peak at 3749 cm^{-1} is owning to a C−H aromatic C−H (aryl) of C−H (aryl) at 5268 cm^{-1} (around 1900 nm) is owning to a O−H str. + 2 × C−O str. of starch. At 6950 cm^{-1} (around 1440 nm) and 7992 cm^{-1} (around 1225 nm) is owning to a 2 × C−H str. + C−H def. and a C−H str. second overtone of CH, respectively. And at 9349 cm^{-1} (around 1065 nm) is owning to a O−H combination band of water. The loading value on PC-3 was found peak at 5222 cm^{-1} (around 1920 nm), it is owing to a C=O str. second overtone of CONH and at 4091 cm^{-1} (2488 nm) is a C−H stretching and C−C stretching combination of cellulose (Otto, 2016). The PCA was re-conducted on those samples. The corresponding score scatter plots is show in (Figure 2a, b, c and d).

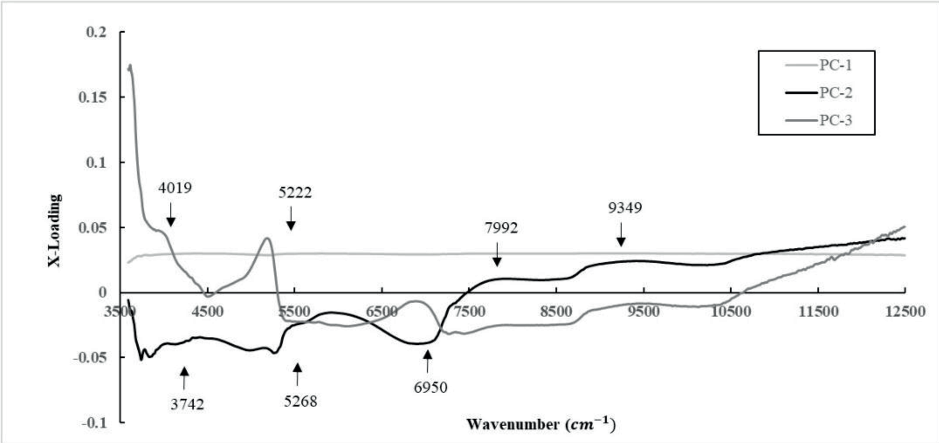

Figure 3. Loading plot for PC-1, PC-2 and PC-3 of stem form

Classification analysis base on raw spectra

Results illustrated in Tables 1, 2 and 3 show the classification of geographical origin with three different methods. In case of SIMCA, the best % accuracy was from intact peel (with thorns) form, with a value of 51.04%. For SVMC, overall classification accuracy achieved high values, providing more than 83.33% CCR. The best % CCR in training and validation set obtained from stem form which was 100% and 100%, respectively. Likewise, for PLS-DA, stem and combined form provided 100% CCR in calibration set though they provided only 45% CCR in the validation set. The best classifying model was provided by SVMC. The separation of overlapping classes is not feasible with methods. SVMC was used because SVMC found optimal separating hyperplanes, it was an efficient solution to separate nonlinear boundaries by constructing a linear boundary in a large, transformed version of the feature space (Otto, 2016).

Table 1. Classification result of geographical origin by SIMCA base on raw spectra

Forms	%Accuracy
Intact peel (with thorns)	*51.04*
Cut-thorns peel	48.94
Stem	25.00
Combine[1]	44.12

[1]Combine = data of all forms (intact peel - with thorns, cut-thorns peel and stem)

Table 2. Classification result of geographical origin by SVMC base on raw spectra

Forms	Set of samples	% CCR[2]
Intact peel (with thorns)	Trainning accuracy	93.75
	Validation accuracy	83.33
Cut-thorns peel	Trainning accuracy	97.50
	Validation accuracy	95.00
Stem	Trainning accuracy	*100.00*
	Validation accuracy	*100.00*
Combine[1]	Trainning accuracy	93.33
	Validation accuracy	91.66

[1]Combine = data of all forms (intact peel - with thorns), cut-thorns peel and stem)
[2]% CCR = % Correct classification rate

Table 3. Classification result of geographical origin by PLS-DA base on raw spectra

Forms	Geographic origin	Calibration set			Validation set		
		No[3]	Missed	% CCR[2]	No[3]	Missed	% CCR[2]
Intact peel (with thorns)	Chan	40	1	97.50	25	20	20.00
	Prachuap	40	0	100.00	23	0	100.00
	Overall classification	80	-	**98.75**	48	-	**60.00**
Cut-thorns peel	Chan	40	0	100.00	23	10	56.56
	Prachuap	40	2	95.00	24	1	95.85
	Overall classification	80	2	**97.50**	47	11	**76.21**
Stem	Chan	20	0	100.00	10	1	90.00
	Prachuap	20	0	100.00	10	10	0.00
	Overall classification	40	0	***100.00***	20	11	***45.00***
Combine[1]	Chan	60	0	100.00	35	30	14.29
	Prachuap	60	0	100.00	33	10	69.70
	Overall classification	120	0	***100.00***	68	40	***42.00***

[1]Combine = data of all forms (intact peel - with thorns, cut-thorns peel and stem)
[2]% CCR = % Correct classification rate
[3]No = number of samples

CONCLUSIONS

This study indicated that NIRs, the high performance method with the advantages of fast analytical speed, easy operation and non-destructive measurement might be a method to guarantee geographic origin of durian Monthong from Prachuap Kiri Khan Province compared with the durian from Chanthaburi Province. However, a more robust model should be developed by using a larger sample group.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, NIRS Research Centre Agriculture Product and Food for supporting equipment and laboratory and the Thailand Research Fund for providing financial during the experiment. Contract number: KMITL-KRIS63-0011

REFERENCES

Arndt, M., Rurik, M., Drees, A., Bigdowski, K., Kohlbacher, O.& Fischer, M. (2020). Comparison of different sample preparation techniques for NIR screening and their influence on the geographical origin determination of almonds (*Prunus dulcis* MILL.). *Food Control*, 115, 107302.

Chanachot, K., Saechua, W. & Sirisomboon, P. (2020). Feasibility of geographical origin discrimination of Monthong intact durian fruits between Trat and Rayong province in the east of Thailand by near infrared spectroscopy. Retrieved from https://www.eng.rmuti.ac.th/ans2020/.

Chen, Y., Xie, M.Y., Yan, Y., Zhu, S.B., Nie, S.P., Li, C., Wang, Y.X. & Gong, X.F. (2008). Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques. *Analytica chimica acta*, 618(2), 121-130.

Duan, X., Zhang, D., Nie, L. & Zang, H. (2014). Rapid discrimination of geographical origin and evaluation of antioxidant activity of *Salvia miltiorrhiza* var. *alba* by Fourier transform near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 122, 751-757.

Eisenstecken, D., Stürz, B., Robatscher, P., Lozano, L., Zanella, A. & Oberhuber, M. (2019). The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. *Postharvest Biology and Technology*, 147, 123-131.

Guo, W., Gu, J., Liu, D. & Shang, L. (2016). Peach variety identification using near-infrared diffuse reflectance spectroscopy. *Computers and Electronics in Agriculture*, 123, 297-303.

Lai, Y., Ni, Y. & Kokot, S. (2011). Discrimination of *Rhizoma corydalis* from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods. *Vibrational Spectroscopy*, 56(2), 154-160.

Lee, D.Y., Kim, S.H., Kim, Y.C., Kim, H.J. & Sung, S.H. (2011). Discrimination of Scrophulariae Radix according to geographical origin and determination of active constituents by near infrared spectroscopy (NIRS). *Microchemical Journal*, 99(2), 213-217.

Luna, A.S., da Silva, A.P., Pinho, J.S., Ferré, J. & Boqué, R. (2013). Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 100, 115-119.

Luo, X., Yu, X., Wu, X., Cheng, Y. & Qu, H. (2008). Rapid determination of Paeoniae Radix using near infrared spectroscopy. *Microchemical Journal*, 90(1), 8-12.

Otto, M. (2016). *Chemometrics: statistics and computer application in analytical chemistry*. John Wiley & Sons.

Osborne, B.G. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation*, https://doi.org/10.1002/9780470027318.a1018.

Sales, E.K. (2015). Durian marker kit for durian (*Durio zibethinus* Murr.) identity. *International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering*, 9(5), 518-528.

Tingting, S., Xiaobo, Z., Jiyong, S., Zhihua, L., Xiaowei, H., Yiwei, X.& Wu, C. (2016). Determination geographical origin and flavonoids content of goji berry using near-infrared spectroscopy and chemometrics. *Food Analytical Methods*, 9(1), 68-79.

Wang, L., Lee, F.S. & Wang, X. (2007). Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content. *LWT-food Science and Technology*, 40(1), 83-88.

Wang, P. & Yu, Z. (2015). Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *Journal of Pharmaceutical Analysis*, 5(5), 277-284.

Woo, Y.A., Kim, H.J., Ze, K.R. & Chung, H. (2005). Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of Angelicae gigantis Radix. *Journal of pharmaceutical and biomedical analysis*, 36(5), 955-959.

Workman Jr, J.&Weyer, L. (2012). *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. CRC press.

Zhu, J.Y., Chen, B., Yan, H., Jia, J.Q. & Han, B.X. (2011). Rapid identification of peucedanum geographical growing areas through near infrared spectroscopy. In *2011 4th International Conference* *on Biomedical Engineering and Informatics (BMEI)*, Vol. 4, 1772-1776, IEEE.

***Bureau of Commercial Provincial Office of the Permanent Secretary Ministry of Commerce (2020). The National Bureau of Agricultural Commodity and Food Standards (ACFS). Retrieved from https://www.acfs.go.th/standard/download/food%20safety%20eng.pdf.

***Camo Software AS, Oslo, Norway. (2020). The unscrambler user manual. Retrieved from https://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Methods.pdf.