การตรวจสอบปริมาณธาตุอาหารหลักในใบทุเรียนพันธุ์หมอนทอง ด้วยเทคนิคเนียร์อินฟราเรดสเปกโทรสโกปี

EVALUATION OF MACRONUTRIENT CONTENT IN DURIAN (CV. MONTHONG) LEAF USING NEAR INFRARED SPECTROSCOPY TECHNIQUE

ฐิติมา พนมโสถณ

THITIMA PHANOMSOPHON

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมเกษตร

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2566

KMITL-2023-EN-D-108-081

EVALUATION OF MACRONUTRIENT CONTENT IN DURIAN (CV. MONTHONG) LEAF USING NEAR INFRARED SPECTROSCOPY TECHNIQUE

THITIMA PHANOMSOPHON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

DOCTOR OF ENGINEERING IN AGRICULTURAL ENGINEERING

SCHOOL OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2023

KMITL-2023-EN-D-108-081

หัวข้อวิทยานิพนธ์	การตรวจสอบปริมาณธาตุอาหารหลักในใบทุเรียนพันธุ์หมอนทองด้วยเทคนิคเนียร์อินฟราเรดสเปกโทรสโกปี
นักศึกษา	นางสาวฐิติมา พนมโสภณ
รหัสประจำตัว	61601168
ปริญญา	วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชา	วิศวกรรมเกษตร
พ.ศ.	2566
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ศ. ดร. ปานมนัส ศิริสมบูรณ์

บทคัดย่อ

ทุเรียนเป็นผลไม้ที่มีมูลค่าสูง และได้รับความนิยมทั้งในประเทศและต่างประเทศ ทุเรียนจึงเป็นพืชเศรษฐกิจที่สำคัญของประเทศไทย ในการปลูกทุเรียนเพื่อให้ได้ผลผลิตที่มีคุณภาพ เกษตรกรมักใส่ปุ๋ยในปริมาณมากเกินความต้องการของต้นทุเรียน จึงเป็นผลให้ต้นทุนการผลิตสูง แต่หากเกษตรกรทราบถึงปริมาณธาตุอาหารในใบ พวกเขาจะสามารถใส่ปุ๋ยได้อย่างเหมาะสม และต้นทุเรียนยังคงผลิตผลได้อย่างมีคุณภาพ พร้อมส่งออกสู่ตลาด วิทยานิพนธ์ฉบับนี้ได้ศึกษาความเป็นไปได้ในการใช้เทคนิคเนียร์อินฟราเรดสเปกโทรสโกปี เพื่อตรวจสอบปริมาณไนโตรเจน (N) ฟอสฟอรัส (P) และโพแทสเซียม (K) ในใบทุเรียนพันธุ์หมอนทอง วัตถุประสงค์ของงานวิจัยแบ่งออกเป็น 2 วัตถุประสงค์หลัก คือ 1) พัฒนาแบบจำลองสำหรับการแบ่งกลุ่มตามระดับความเข้มข้นของธาตุอาหาร โดยสร้างแบบจำลองจากสเปกตรัมของใบสด และใบบดแห้ง สแกนด้วย FT-NIR spectrometer จำนวนคลื่น 12500-4000 cm^{-1} (ความยาวคลื่น 800-2500 nm) และวิเคราะห์ค่าความเข้มข้นของ N วิเคราะห์ด้วยเครื่อง TruMac CNS ส่วน P และ K วิเคราะห์ด้วยเครื่อง ICP-OES โดยจัดเตรียมข้อมูลให้แต่ละกลุ่มมีจำนวนตัวอย่างใกล้เคียงกันจากการสังเคราะห์ข้อมูลด้วยวิธี synthetic minority oversampling technique (SMOTE) เพื่อลดปัญหาการละเลยการตรวจสอบกลุ่มตัวอย่างที่มีจำนวนประชากรน้อย แบบจำลองที่ดีที่สุดถูกพัฒนาด้วยวิธีโครงข่ายประสาทเทียม พบว่าค่าความแม่นยำในการแบ่งกลุ่มจากใบสดไม่แตกต่างจากใบบด จึงเลือกใช้แบบจำลองจากใบสด โดยมีค่าความแม่นยำในการแบ่งระดับความเข้มข้น N P และ K เท่ากับ 0.99 0.97

และ 1.00 ตามลำดับ 2) พัฒนาแบบจำลองสำหรับการวัดปริมาณความเข้มข้นของธาตุอาหาร โดยสร้างแบบจำลองจากสเปกตรัมของใบสด ใบบดแห้ง และใบบดแห้งอัดเม็ด และใช้วิธีการสแกนและวิเคราะห์ค่ามาตรฐานด้วยวิธีเดียวกัน แบบจำลองถูกพัฒนาด้วยสมการการถดถอยกำลังสองน้อยที่สุดบางส่วน และสร้างแบบจำลองจากช่วงคลื่นต่างๆ ได้แก่ ช่วงคลื่นแบบเต็มช่วง แบบช่วงสั้น และแบบเลือกเฉพาะบางช่วงคลื่นด้วยวิธี successive projections algorithm (SPA) เมื่อเปรียบเทียบแบบจำลองจากการเลือกใช้ช่วงคลื่นต่างๆ พบว่าทุกแบบจำลองให้ค่าสัมประสิทธิ์การพิจารณาไม่แตกต่างกัน จึงแนะนำให้ใช้แบบจำลองที่สร้างจากเฉพาะบางช่วงคลื่น เพื่อความรวดเร็ว โดยค่าสัมประสิทธิ์การพิจารณาของชุดทำนายที่ดีที่สุดสำหรับการทำนายปริมาณความเข้มข้นของ N ในใบสด เท่ากับ 0.852 และสำหรับ K ในใบบดแห้ง เท่ากับ 0.834 เมื่อพิจารณาเกณฑ์ค่าสัมประสิทธิ์การพิจารณาแบบจำลองเหล่านี้สามารถทำนายได้ในระดับดี แต่สำหรับ P ทุกแบบจำลองอยู่ในเกณฑ์ที่ไม่สามารถนำไปใช้ได้ เนื่องจากไม่พบความสัมพันธ์ระหว่างสปกตรัมกับ P ในใบทุเรียน จากผลการวิจัยนี้สนับสนุนว่าเทคนิคเนียร์อินฟราเรดเป็นเครื่องมือที่สามารถแบ่งกลุ่มตามระดับความเข้มข้นของ N P และ K ในใบทุเรียน และวัดปริมาณความเข้มข้นของ N และ K ในใบทุเรียนได้อย่างรวดเร็ว เมื่อเทียบกับวิธีมาตรฐานดั้งเดิม

คำสำคัญ: เนียร์อินฟราเรดสเปกโทรสโกปี, ใบทุเรียน, ไนโตรเจน, ฟอสฟอรัส, โพแทสเซียม

Thesis	EVALUATION OF MACRONUTRIENT CONTENT IN DURIAN (CV. MONTHONG) LEAF USING NEAR INFRARED SPECTROSCOPY TECHNIQUE
Student	Miss Thitima Phanomsophon
Student ID.	61601168
Degree	Doctor of Engineering
Program	Agricultral Engineering
Year	2023
Thesis Advisor	Prof. Dr. Panmanas Sirisomboon

ABSTRACT

Durian is a highly valuable fruit that is popular both domestically and internationally. It is an important economic crop in Thailand. However, in the cultivation of durian, farmers tend to apply more fertilizer than necessary to produce high-quality fruits. This results in high production costs. If farmers know the nutrients content in the leaves, they will be able to apply fertilizer more accurately, allowing the durian trees to produce quality fruits while reducing costs. This thesis investigates the feasibility of using near infrared spectroscopy (NIRs) to determine the nitrogen (N), phosphorus (P), and potassium (K) content in durian (CV. Monthong) leaf. The research objectives were divided into two main goals. 1) to develop a model for classification nutrient content levels. In this study, spectra of fresh and dried ground leaves were collected using an FT-NIR spectrometer with wavenumber from 12500-4000 cm^{-1} (wavelengths from 800-2500 nm). The nutrients content was analyzed by TruMac CNS analyzer for N, and ICP-OES analyzer for P and K. Additional data synthesis by Synthetic minority oversampling technique (SMOTE) was used to balance the sample groups and mitigate the problem of neglecting

predictions for minority class samples. The artificial neural networks (ANN) algorithm yielded the best model. The accuracy of the models for fresh and dried ground leaves did not differ significantly, thus recommending the fresh leaves predicting. The achieving accuracy in classifying N, P and K content levels, with accuracies of 0.99, 0.97, and 1.00 respectively. 2) to create a model for nutrient context predicting with partial least squares regression equations. Spectra of fresh, dried ground, and pellet leaves were scanned and analyzed same method with objective 1, and the model was developed by full wavelength, short wavelength, and specific wavelengths that were selected by the successive projections algorithm (SPA) which was observed that each wavelength produced similar r^2. Therefore, it is recommended to use a model built with SPA for quick measurement. The r^2 of the best model for predicting N content in fresh leaves was 0.852, while for predicting K content in dried ground leaves, the r^2 was 0.834. These models demonstrated good predictive capability based on the coefficient criteria. However, for predicting P content, none of the models met the criteria because of non-relationship between NIR spectra and P in durian leaf. The results of this research support the use of NIRs as a tool for detecting N, P, and K content levels in durian leaves. And it enables the rapid quantification of N and K content compared to traditional methods.

Key words: near-infrared spectroscopy, durian leaf, nitrogen (N), phosphorus (P), potassium (K)

Acknowledgements

I wish to express my deepest and sincere gratitude to my advisor, Professor Dr. Panmanas Sirisomboon, for her valuable supervision, advice, support, immense knowledge, and encouragement throughout the course of my study. I would also like to express my thanks to Assistant Professor Dr. Nukoon Tawinteung, Assistant Professor Dr. Lampan Khurnpoon, Associate Professor Dr. Jetsadapon Posom, Assistant Professor Dr. Jiraporn Sripinyowanich Jongyingcharoen, and Assistant Professor Dr. Ravipat Lapcharoensuk for their helpful recommendations. I am grateful to Mr. Natthapon Jaisue, Mr. Akarawhat Worphet, and Mr. Bijendra Shrestha for their assistance in conducting the experiment.

I am particularly grateful for the financial support provided by the Royal Golden Jubilee scholarship (RGJ) PhD program [Grant numbers NRCT5-RGJ63021-169] from the National Research Council of Thailand (NRCT), which covered tuition fees, research expenses, monthly expenses, and international research experience. I would also like to express my gratitude to the Agricultural Research Development Agency (Public Organization) [grant number PRP6305031290], which covered the research expenses.

Furthermore, I would like to express my sincere gratitude to Professor Dr. Satoru Tsuchikawa, Associate Professor Dr. Tetsuya Inagaki, and Assistant Professor Dr. Te Ma for allowing me to join their laboratory at Nagoya University, Japan, and for their invaluable help, constant encouragement, and immense knowledge throughout the international research experience.

I would like to thank the Near Infrared Spectroscopy Research Center for Agricultural Products and Food (www.nirsresearch.com) at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, and the Graduate School of Bio Agricultural Sciences, Nagoya University, Nagoya, Japan, for providing laboratory space and instruments. Additionally, I am grateful to the Department of Plant Production Technology,

Faculty of Agricultural Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, for providing leaf samples and nutrient analysis.

My heartfelt appreciation goes to all the faculties of the Department of Agricultural Engineering at King Mongkut's Institute of Technology Ladkrabang (KMITL), who not only taught me research methodologies but also supported me in all aspects of life, providing motivation and support. I would also like to thank my friends for their wonderful collaboration, as they greatly supported me during the experiments and were always willing to help.

Finally, I would like to express my appreciation to my family for their love and support, both physically and mentally, throughout the course of my study. Without them, I cannot imagine my success today. Thank you so much to my father and my mother, you mean everything to me. I consider myself fortunate to have you in my life.

Thitima Phanomsophon

Table of contents

Page

Table of contents (continued)

Page

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

Page

List of tables

List of tables (continued)

List of figures

List of figures (continued)

Abbreviations

1D	1st derivative
2D	2nd derivative
ANN	Artificial neural networks
Bias	Average error
Cal	calibration set
DL	Dried ground leaves
E	Misclassification
FL	Fresh leaf
FT-NIR	Fourier transform near-infrared
K	Potassium
kNN	k-nearest neighbour
LOQ	Lowest limit of quantification
MC	Mean centring
MMN	Min-max normalization
MN	Mean normalisation
MSC	Multiple scatter correlation
N	Nitrogen
NIR	Near Infrared
nLVs	Number of latent variables (PLS factors)
nWL	Number of wavelengths
P	Phosphorus
PC	Principal components
PCA	Principal component analysis
PL	Dried ground leaves pellet
PLS	Partial least squares regression

Abbreviations (continued)

R^2	Coefficient of determination
Rep	Repeatability
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
RPD	Ratio of prediction to deviation
SD	Standard deviation
SECV	Standard deviation of error of cross validation
SEP	Standard deviation of error of prediction
SMOTE	Synthetic minority oversampling technique
SMT	Smoothing
SNV	Standard normal Variate
SPA	Successive projections algorithm
THz	Terahertz
TP	true positives
Unk	Unknown set
Val	Validation set

Chapter 1

Introduction

1.1 Statement and significance of the problems

Durian is known as the king of fruits, thanks to its powerful and sweet aroma that makes it incredibly tasty. It also holds a significant place in Thailand's economy as the major fruit export. According to durian export data, the quantity has been increasing each year, especially in the last two years (2021-2022), where there has been a 30-40% increase in export quantity compared to 2020. Furthermore, the value has risen by 66-68% [1]. These statistics indicate the growth of the Thai durian export market, leading to more farmers choosing to cultivate durian in Thailand's southern, eastern, and northeastern regions. To ensure the production of high-quality durian fruit, farmers need to take care of their durian trees by providing appropriate watering and fertilization. This is why farmers often apply more fertilizer than necessary, which increases production costs. However, it is undeniable that fertilization is essential for plants. If we compare trees to people, nutrients are like food that enables proper growth and development. To survive and thrive, plants require essential nutrients for growth, development, and grain production. Each nutrient serves a specific function that cannot be entirely replaced by any other nutrient. Insufficient nutrients can hinder growth, while an excess of nutrients can even result in plant death [2]. Therefore, assessing the nutrient content in the leaves is necessary to determine the appropriate amount of fertilizer, allowing farmers to save costs and increase revenue.

Traditionally, chemical methods have been used to determine the nutrient content in leaves, but these methods are time-consuming. Nowadays, researchers utilize

the near infrared (NIR) technique as a rapid and non-destructive method to analyze nutrient content in leaves instead of using chemical methods [3]. For example, Azadnia et al. conducted a study on N, P, and K content in apple leaves [4], Guo et al. focused on N and P content in tea leaves [5], Rébufa et al. explored N and K content in Moringa oleifera leaves [6], Rotbart et al. investigated N and K content in olive leaves [7], Yarce and Rojas examined N, P, and K content in sugarcane [8], Guo et al. studied P content in rubber leaves [9], and Wang et al. investigated P and K content in tea leaves [10], among others. However, there seems to be no existing study on the use of NIR spectroscopy to determine the nutritional content of durian leaves.

If farmers can obtain timely information about the nutrient levels in durian leaves, they can make informed decisions regarding fertilization. This not only helps reduce costs for farmers but also contributes to saving the environment by reducing resource wastage. Supporting sustainable agriculture is essential for the well-being of our planet.

1.2 Goal and objective

The overall objective of this thesis was to investigate the possibility of using NIR techniques to evaluate the nutrient content (specifically nitrogen (N), phosphorus (P), and potassium (K)) in *Durio zibethinus* Murray CV Monthong leaves, including fresh leaves, dried ground leaves, and dried ground leaves pellets. The research had three specific objectives, outlined as follows:

1) To identify a suitable algorithm for developing a classification model to classify nutrient levels in fresh durian leaves and dried ground leaves using near infrared spectroscopy.
2) To examine the impact of imbalanced data on the classification model's accuracy of fresh durian leaves and dried ground leaves using near infrared spectroscopy.

3) To develop a regression model using near infrared spectroscopy to assess the nutrient content in fresh leaves, dried ground leaves, and dried ground leaf pellets.
4) To investigate methods for reducing the number of wavelengths that affect the performance of the NIR regression model in fresh leaves, dried ground leaves, and dried ground leaf pellets.
5) To explore the application of Terahertz spectroscopy as a new technique for assessing the nitrogen content in dried ground leaf pellets.

1.3 Hypothesis to be tested

The hypothesis is that the near-infrared spectroscopy technique can be used as a reliable and accurate method for evaluating the macronutrient content in durian leaves.

1.4 Scope or limitation of the study

In this thesis, *Durio zibethinus* Murray CV Monthong leaves from trees older than eight years at orchards in Rayong, Chanthaburi, and Trad provinces in Thailand's eastern region were used to conduct a preliminary study on the possibility of applying NIR spectroscopy to determine their levels or content. Different sample metrics, including fresh leaves, dried ground leaves, and dried ground leaf pellet, were compared to identify the most suitable sample type for prediction.

For the spectral measurements, an FT-NIR spectrometer (MPA, Bruker Ltd., Germany) with a wavelength range of 12,500-4,000 cm^{-1} (800-2,500 nm), average of 32 scan, a resolution of 32 cm^{-1}, gold as the reference material and under laboratory conditions at 25±2°C was used to determine the levels or content. The fresh leaf was scanned under an aluminum plate in 2 scan positions. The spectra of one fresh leaf were averaged from 20 fresh leaves. The dried ground leaf sample was scanned within a quartz

cup (diameter 4.5 cm and height of 5.3 cm). The dried ground leaf pellet sample was covered with Spectralon reference material and covered with an aluminum can before scanning. The Tera Prospector (Nippo, Precision Co., Ltd., Japan) with a horizontally polarized beam, a bandwidth of 0.1-4.0 THz, an average of 100 scans and the spectral resolution was 0.02 THz, air as the reference material, relative humidity in THz optical system at 0.0-0.1% and under room conditions at 26°C was also employed for measurements for terahertz spectra.

For the nutritional analysis, CNS analyzer (Leco, USA) was utilized for nitrogen (N) analysis, while ICP-OES (inductively coupled plasma optical emission spectrometer) (Perkin Elmer, USA). was applied for phosphorus (P) and potassium (K) measurement. These instruments are used to measure the nutrient content in the standard method (AOAC 984.27 for P and K analysis).

1.5 Process of the study

This thesis is organized into eight chapters, each covering specific aspects related to the research topic, as follows:

Chapter 1 presents the research background, problem statement, and research objectives of this thesis.

Chapter 2 provides detailed information on durian, nutrients, nutrient analyzers, NIR spectroscopy, and model development using partial least squares regression, k-nearest neighbor, and artificial neural networks. It also includes literature reviews on the applications of NIR spectroscopy for nutrient determination in leaves.

Chapter 3 illustrates the precision test of NIR scanning for durian leaves (fresh leaves, dried ground leaves, and dried ground leaf pellets) and analyzes the precision of reference laboratories for nutrient analysis. The purpose of this chapter is to ensure the accuracy

and suitability of NIR spectral recordings from FT-NIR and the standard reference of nutrient content for model creation.

Chapter 4 focuses on the classification of N, P, and K concentration levels using the FT-NIR spectrometer in fresh leaves and dried ground leaves of durian. It also addresses the effects of imbalanced data on modeling. The purpose of this chapter is to compare algorithms and sample types for classification modeling and to improve imbalanced data for increased model accuracy. The chapter concludes with the construction of the nutrient level model.

Chapter 5 presents the regression of N, P, and K concentration using the FT-NIR spectrometer in fresh leaves, dried ground leaves, and dried ground leaf pellets of durian. It also discusses the effects of wavelength selection on model performance. The purpose of this chapter is to compare sample types for regression modeling and to select suitable wavelengths for modeling. The chapter concludes with the construction of the nutrient content model.

Chapter 6 focuses on determining the lowest quantification of the regression model for N, P, and K concentration using the FT-NIR spectrometer in fresh leaves, dried ground leaves, and dried ground leaf pellets of durian.

Chapter 7 explores the evaluation of nitrogen in durian leaves using Terahertz time-domain spectroscopy. The purpose of this chapter is to investigate a new technique for nitrogen determination, and the results are presented.

Chapter 8 presents the overall main conclusions and recommendations for further study based on the findings from chapters 3, 4, 5, 6, and 7 of this thesis.

1.6 Reference

[1] Office of Agricultural Economics. "Durian Export Statistics" [Online]. Available: https://www.oae.go.th/view. 2023.

[2] Aftab, T. and Hakeem, K. R., "Plant nutrition and soil fertility: physiological and molecular avenues for crop improvement," **Sustainable Plant Nutrition**: Elsevier Inc., 2022.

[3] Prananto, J. A., Minasny, B., and Weaver, T., "Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues," **Advances in Agronomy**: Elsevier Inc., 2022.

[4] Azadnia, R., Rajabipour, A., Jamshidi, B., and Omid, M. "New approach for rapid estimation of leaf nitrogen, phosphorus, and potassium contents in apple-trees using Vis/NIR spectroscopy based on wavelength selection coupled with machine learning" **Computers and Electronics in Agriculture**, vol. 207. 2023. https://doi.org/10.1016/j.compag.2023.107746.

[5] Guo, J., Huang, H., He, X., Cai, J., Zeng, Z., Ma, C., Lü, E., Shen, Q., and Liu, Y. "Improving the detection accuracy of the nitrogen content of fresh tea leaves by combining FT-NIR with moisture removal method" **Food Chemistry**, vol. 405. 2023. https://doi.org/10.1016/j.foodchem.2022.134905.

[6] Rebufa, C., Pany, I., and Bombarda, I. "NIR spectroscopy for the quality control of Moringa oleifera (Lam.) leaf powders: Prediction of minerals, protein and moisture contents" **Food Chemistry**, vol. 261. 2018. pp.311-321. https://doi.org/10.1016/j.foodchem.2018.04.066.

[7] Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., and Yermiyahu, U. "Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance" **Biosystems Engineering**, vol. 114. 2013. pp.426-434. https://doi.org/10.1016/j.biosystemseng.2012.09.005.

[8] Yarce, C. J. and Rojas, G. "Near infrared spectroscopy for the analysis of macro and micro nutrients in sugarcane leaves" **Sugar Industry**, vol. 2012. pp.707-710. https://doi.org/10.36961/si13611.

[9] Guo, P. T., Zhu, A. X., Cha, Z. Z., Li, M. F., and Luo, W. "A local model based on environmental variables clustering for estimating foliar phosphorus of rubber trees with vis-NIR spectroscopic data" **Heliyon**, vol. 8. 2022. pp.e09795. https://doi.org/10.1016/j.heliyon.2022.e09795.

[10] Wang, Y.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., and Zhang, Z.-Z. "NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves" **Infrared Physics & Technology**, vol. 108. 2020. https://doi.org/10.1016/j.infrared.2020.103365.

Chapter 2

Literature review

2.1 Overview of Durian

The king of fruits, durian, is a famous tropical fruit that originated in Southeast Asia. It is a fruit with a unique smell and sweet taste that is popular in Thailand and all over world, particularly in China, Hong Kong, Taiwan, and Singapore. Furthermore, durian fruit has a rich nutrition in terms of fat, protein, carbohydrates, and minerals. Thailand is a major supplier of durian because Thai durian smells and tastes delicious. Due to its rising export every year [1], durian is regarded an important commercial crop in Thailand. It is also a costly fruit that is marketed widely. Durian planting is popular among farmers, and it can also be processed to improve the value of commodities such as durian frizz dry, fried durian, or durian stirrup, among others. As a result, processed durian is popular for international tourists who purchase it as a souvenir. Nonetheless, fresh durian remains the most popular.

2.1.1 Characteristics of durian [2]

1. **The trunk** is a large perennial plant with softwood, ranging from 6 to 24 inches in diameter, a height of 70 to 80 feet, and a lifespan of 80 to 150 years. The skin is rough and fissured, with a grayish tone. The trunk's longitudinal branches are long and rounded. Depending on the type of durian and the amount of red-light exposure, the branches can be straight or curled. The top of the durian can be classified into three types: square,

Figure 2.1 Monthong durian tree.

inverted cone with a wide base, and inverted cone with a narrow base, based on the characteristics of durian varieties.

2. The leaves are single-leaved, broad, and cotyledon, measuring 2 to 3 inches in width and 6 to 8 inches in length with a pointed tip. The leaves are dark green when mature, with brown undersides, and they sprout from the buds on the branches.

Figure 2.2 Durian leaves.

3. The roots forage along the soil surface up to a depth of 50 centimeters. Durian has adventitious roots that absorb water and nutrients. It lacks hairy roots. The taproot of the durian tree serves to anchor the trunk, while lateral roots and fibrous roots aid in foraging and anchoring the stems.

4. flowers resemble bell shapes. They are perfect flowers with greenish-brown sepals (outermost petals), five soft white petals, stamens (five sets of 5 to 8 filaments, each with numerous pollen sacs), and a stigma (with a central style). Durian flowers often bloom in clusters, with a bouquet containing 1 to 30 flowers.

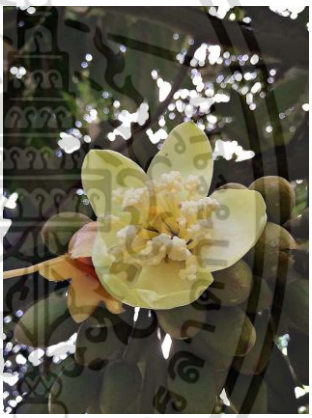


Figure 2.3 Durian flowers [3].

Figure 2.4 Monthong durian fruit [4].

5. The fruits is capsule-based. The skin is covered in hard spikes and ranges in color from dark green to yellowish-green. Durian fruits come in various shapes such as round, oval, and obtuse bottoms. They have

a diameter of approximately 15 to 20 cm and a length of 25 to 35 cm. Typically, each fruit has 5 to 6 lobes, and each lobe contains a single seed. The pulp is white and can have shades of yellow, light yellow, orange, and more.

2.1.2 Growth stage [5-9]

2.1.2.1 The unproductive stage (the first 1-4 years)

This stage is the initial stage of durian planting. The young durian trees are particularly fragile, especially at the base and roots. So, special maintenance and care are necessary, as they are constantly fallen root, worm borer, and root rot. The durian must be watered immediately after planting and every 3 days, around 5 liters at a time. When the durian is 8-10 months old, the soil is shoveled, and the hay is covered with grass so that the roots do not get too much sun, dry up and die. Young leaves will sprout when the durian is a year old. The farmer should cultivate the soil at a larger distance according to the durian bush area, apply chemical fertilizer alternately with manure, and cover the grass around the roots. If roots spring up from the earth, cover them with soil to prevent them from drying out and dying. When the durian is 2 years old, the durian will settle and grow well. The farmer should water every 3 days, cultivate the soil, apply fertilizer, and cover the grass as usual. The durian can be leaf trimmed when it is three years old. Flowers may bloom at the end of the year, but the farmer must chop them off since the roots of the durian tree are not yet strong enough, causing the durian tree to stop growing. The flowers begin to blossom more in the fourth year. However, farmers should allow just 4-5 blooms for fruiting to avoid over-deterioration of the durian tree and should apply more fertilizer.

Figure 2.5 Durian seedlings [10].

In the term, it also does not give some fruit, the durian trees need more nitrogen (N) than other trace elements. Due to N is a nutrient that helps plants grow well, strong trunk and branching. The fertilizer (Nitrogen-Phosphorus-Potassium) used should depend on the type of soil as follows: 1) clay soil, use fertilizer 20-10-10 amount 500 g per 1 m diameter of the canopy, 2) loam soil, use fertilizer 15-15-15 amount 700 g per 1 m diameter of the canopy, 3) sandy soil, use fertilizer 20-10-10 amount 1,000 g per 1 m diameter of the canopy, divided into fertilizer 4 times a year, every 3 months. In addition, organic fertilizer should be applied with about 10-20 kg per tree. However, it is recommended to trim only two generations of flowers: probably cutting the 1st and 3rd generations to only the 2nd generation or reducing the 2nd generation flowers to only the 1st and 3rd generations.

2.1.2.2 The flowering stage

A durian flower group which will bloom under the branches. Durians grown with seeds begin to bloom at the age of 7-8 years, but durians grown with rootstocks will bloom in the 4th year, or perhaps sooner if the soil is good. In each flower group there are about 25 flowers, but only 2-3 fruiting flowers are left. Normally, durians flower has about 2-3 generations a year, each generation has a gap of 20-50 days.

Durian flower growth stage is divided into 6 stages [11]:

1 **Roe phase:** The new flowers begin to bloom which is small blister like the roe with light-brown in color.

2 **Rat's foot stretching phase:** The small blister became to a small flower sticking out of the branch.

3 **Button phase:** The flowers grow, they long protrude like pendulums or buttons.

4 **Bracelet head phase:** The flowers grow rapidly, both wide and long. The pollen inside the flower begins to push outwards, resembling a child's ankle bracelet head.

5 **Flowering phase:** The stamen and pistil are expand and push out. The sepals will spring into 2 petals.

6 **Clasp or pin phase:** When the flowers are in full bloom, petals will begin to fall off until only the pistil remains. It is like a clasp or a pin.

Figure 2.6 Roe phase [12].

Figure 2.7 Rat's foot stretching phase [12].

Figure 2.8 Button phase [12].

Figure 2.9 Bracelet head phase [12].

Figure 2.10 Flowering phase [12].

Figure 2.11 Clasp or pin phase [12].

At the flowering stage, durian plants need higher phosphorus (P) and potassium (K), as this will help with more flowering and strong flowering. Fertilizer 9-24-24 should be

used. If farmers want to speed up flowering, use fertilizer 10-52-17 amount 100-200 g per 1 m diameter of the canopy, shoveling the soil, fertilizing and watering.

2.1.2.3 The fruiting stage

Fruiting is the transition from flowering to immature fruit that proceeds following successful pollination. Durian flowers bloom in the late afternoon from 1-2 p.m., the stigma will be ready to pollinate, and the stamens will begin to discharge anthers in the evening at 4-5 p.m. The petals and stigma will bend in from 6-7 p.m. after full flowering. Pollen will be released from the stamens and mixed with the stigma. The best period for pollination is between 8 and 10 p.m. When pollination is successful, the stamens fall down the style to the ovules, where they release the genetic material into the eggs. Fertilization induces physiological changes in the eggs, and cell division results in immature fruit.

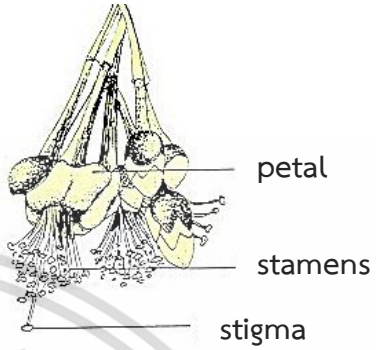


Figure 2.12 components of durian flower.

The development of durian fruit is separated into 4 phases [13]:

1 **0-2 weeks after flowering:** The fruit is rather tiny. Tissues and organs will develop slowly. Because there is so much fruit, competition amongst the fruits causes the nutrients drawn from the stems to be relatively low. The perfect and well-positioned fruits will be able to obtain enough food for growth, whereas other fruits will eventually fade away since they do not get enough food.

2 **3-7 weeks after flowering:** Tissues and organs will develop quickly, especially in the bark. As a result, there is an increase in nutrient drawing, as well as a durian fallout effect due to nutritional deficiency. However, it is dependent on the supply of the food; if the durian tree has sufficient food

sources to fulfill the requirements of each fruit, the amount of fruit shedding can decrease.

3 **8-12 weeks after flowering:** Tissues and organs grow at a breakneck pace. Because seeds and durian pulp grow quickly, amounts of food are drawn up. If durian is insufficient food supply to fulfill the demand, resulting in a tiny, non-growing, and disproportionately shaped durian impact. If the condition is severe, the fruit will be discarded, and the pulp effect may result in uneven flesh color.

4 **13-16 weeks after flowering:** Durian fruit will be completely ripe and ready for harvest. If the durian tree has enough food sources, it produces a healthy yield with standard size, numerous, and high quality. Nevertheless, if the durian tree does not have enough food sources, the resultant output is tiny, late ripening, and of low quality.

Durian trees require more potassium (K) during fruiting because it increases the quality of the fruit. If the food supply from the durian tree is insufficient, the weakest impact may be lost, and the durian fruit may be of poor quality. As a result, adequate fertilizer and water must be given. The fertilizer 13-13-21 should be applied at a rate of 85-100 g per 1 m diameter of the canopy. The sulfur-soluble may be utilized, but not in overabundance since it may be damaging to the durian tree.

2.1.2.4 The harvesting stage

When the durian fruit is mature, it is ready to be harvested. The durian fruit can be divided into 3 types according to the age of the fruit [13].

1 **Light weight:** The harvesting period is 95-105 days after flowering, such as durian CV Luang, Kradum Thong, Chani, Khiaosaat, etc.

2 **Middle weight:** The harvesting period is 105-120 days after flowering, such as durian CV Kan Yao, Chomphusi, Thongyoichat, Monthong, etc.

3 **Heavy weight:** The harvesting period is 120-140 days after flowering, such as durian CV Chaimafai, etc.

Only mature durian fruits should be harvested for good output. Due to a variety of causes, not all durian fruits mature at the same time. As a result, the following techniques for determining durian fruit that is ready to has been used harvested [13]:

1 **Observe the fruit stem:** The fruit stem must be in hard and dark color. A pestle is felt while touching the stalk. It feels like spring as the stem swings.
2 **Observe the thorns:** The thorn tips are dark brown, dry, and easily broken. It feels like spring when squeeze the thorns together.
3 **Observe the slit line between the lobes:** It will be able to clearly observe the slit between the lobes.
4 **Stem tasting:** When the fruit stem is cut off, clear liquid appears. It should not gooey and has a sweet taste.
5 **Thorns knocking:** When knocking on the thorns, there will be a squeaky sound.
6 **Durian fruit fall:** When durian fruit begins to fall, it indicates that the durian is mature, old, and ready for harvest.
7 **Age counting:** It is countable after flowering according to the harvesting period.

Farmers must maintain durian trees after harvesting to a collect enough food to bear fruit for the next fruitful season, which can be done as follows [13]:

1 **Before the completion of harvesting:** Pruning the tree that has already harvested fruits to hasten the sprouting of new leaves which it will increase air circulation inside the canopy, prevent disease and insect spreading, and raise carbon dioxide within the canopy. As a result, the durian tree will effectively photosynthesize to collect food for the next season.

2 **Post-harvest:** To compensate for the lost nutrients, the farmer should apply fertilizer 15-15-15 around the canopy at a rate of 350 g per 1 m diameter of the canopy, and it should be mixed with organic fertilizer at a rate of 10-20 kg per plant.

3 **Sprouting of new leaves:** The young leaves must be fully preserved by spraying chemicals to prevent diseases and insects, and foliar nutrients must be provided. The farmer should apply 300 g of fertilizer 7-13-34 + 12.5 zinc per plant.

4 **Before flowering:** To accelerate flowering, the farmer should prune and treat with 9-24-24 or 12-24-12 fertilizer in amounts of 2-5 kg per plant.

2.1.3 The nutrient content of durian leaves

The proper nutritional content in durian leaves which helps durian trees produce quality fruit was shown in Table 2.1. The durian tree obtains sufficient nutrients, which not only improves its ability to produce good fruit. It also helps the plant to be strong, complete, to produce a lot of blossoms, and to be ready to produce fruit in the next season.

Table 2.1 The appropriate nutrient content of durian leaves [14].

Nutrients	Appropriate content
N - Nitrogen (%)	2.00 - 2.25
P - Phosphorus (%)	0.15 - 0.25
K - potassium (%)	1.00 - 2.00
Ca - Calcium (%)	0.90 - 1.20
Mg - Magnesium (%)	0.40 - 0.60
Fe - Iron (ppm)	40.00 - 60.00
Mn - manganese (ppm)	40.00 - 200.00
Zn - zinc (ppm)	15.00 - 25.00

2.2 Primary nutrients

Nitrogen (N), phosphorus (P), and potassium (K) are primary nutrients that plants require in substantial quantities than other nutrients. It assists growth of plants and the completion of the plant life cycle, as well as a lack of it, can have an influence on crop production and growth. The cultivation of durians is similar. Durian plants require proper nutrition that will provide good and high-quality outcomes. Each primary nutrient is important to plant as follows [15]:

2.2.1 Nitrogen (N)

Nitrogen is a non-metal element with atomic number 7 and mass 14.008u. Approximately 78% of the nitrogen in the atmosphere is in the gaseous state, which plants cannot directly take, but may absorb in the form of ions such as nitrate ions (NO_3^-) and ammonium ions (NH_4^-), and molecules such as urea [$CO(NH_2)$] and amino acids. Nitrogen is essential in the following:

Activation of plant growth: nitrates act as signals which regulate metabolism and the physiological development of plants such as root development, leaf development, seed germination and flowering.

Protein composition: nitrogen is the most abundant element in proteins. Proteins are structural components of the cytoplasm, tissues, and enzymes in plants.

Amino acids composition: nitrogen is a component of glutamic acid that photosynthesis occurs. The amino acids are components of proteins that send a signal to the ends of roots for growth.

Plant hormones composition: nitrogen is included in the auxin hormone and cytokinin hormone. Auxin hormone is a growth hormone that stimulates cell division, accelerates cell proliferation, regulates roots, and prevents leaf, branch, and fruit fall. Cytokinin hormone is responsible for increasing cell division, cell

growth buds, seed germination, and protein production, as well as decreasing leaf stagnation and assisting in nutrition movement.

Nucleic acids composition: nitrogen is a component of nucleic acids, which nucleic acids relate to protein synthesis and genetic storage.

Ammonium composition: nitrogen is a component of ammonium, which is utilized in the production of fertilizer. However, if the plant absorbs too much ammonium, its development would be slowed.

Development of seedlings: nitrogen influences photosynthesis and seedling growth because plants use nutrients from endosperm and convert cotyledons for photosynthesis during the early stages of germination. Plants that do not obtain adequate nitrogen, on the other hand, have a poor rate of photosynthesis and sluggish development.

For durian planting, nitrogen promotes plant growth and strength. This includes assisting in the nourishment and repair of damaged organs. To be ready for durian production [14].

2.2.2 Phosphorus (P)

Phosphorus is a non-metal element with atomic number 15 and mass 30.975u. This element interacts readily with oxygen. As a result, phosphorus does not occur as an element, but rather as an inorganic phosphate such as calcium phosphate ($Ca_3(PO_4)_2$). Phosphorus has a lower concentration of macronutrients in plants than nitrogen and potassium. Phosphorus is essential in the following:

Nucleic acids composition: phosphorus is an important component in nucleic acids which is the structure of the DNA strands used to store genetic information.

Phospholipids composition: phosphorus is a structural component of phospholipids in membranes. Phospholipids help to keep cell membranes stable.

Coenzymes composition: phosphorus is a component in coenzymes such as NAD+ (nicotinamide adenine dinucleotide) that used in the respiration process, NADP+ (nicotinamide adenine dinucleotide phosphate) that used in photosynthesis, FAD (Flavin adenine dinucleotide) that help catalyzes oxidation-reduction reactions, and coenzyme A that used in the decomposition and synthesis reactions of lipids.

Phytate composition: phosphorus is a component in phytate, which is a component of seeds, pollen, and roots, and regulates starch synthesis in seeds.

Metabolism regulation: phosphates regulate carbohydrate metabolism and photosynthesis.

Myoinositol signals: plants use myoinositol phosphates for hormone signaling. Phosphatidylinositol phosphates are used as signal regulators from plant disease.

For durian planting, phosphorus aids in the acceleration of flowering, resulting in an abundance of flowers on the durian tree [14].

2.2.3 Potassium (K)

Potassium is an alkali metal element with atomic number 19 and a mass 39.10 u. It easily reacts with oxygen in both air and water. Potassium is absorbed by plants as potassium ions (K^+). Potassium is easily transported both within cell and cell-cell in tissues of plants. Potassium is essential in the following:

Electrical control: potassium controls the electrical of inorganic and negatively charged organic matter such as anions of organic acids, DNA and phospholipids and controls the electric potential of the membrane.

Enzyme activation: potassium serves to stimulate the activity of about 50 enzymes.

Protein synthesis: potassium is essential in the protein synthesis of multi-cellular plants. The optimum rate of ribosome protein production is dependent on proper potassium consumption.

Photosynthesis: potassium encourages the synthesis of ATP (adenosine triphosphate) in photophosphorylation process and maintains the structure of chloroplast and piroplasmids suitable for carbon dioxide fixation.

Cell enlargement: before cell enlargement, potassium helps to stabilize the pH in the cytoplasm and reduce the osmotic potential in the vacuole. If plants have enough potassium in combination with cytokinin, they can help grow cells four times more well than potassium-deficient plants.

Opening and closing of the stomata: when the potassium concentration in the control cells rises, water transfers from the apoplast into the control cell, causing the cells to turgor pressure and the stomata to open; when the potassium concentration in the control cells drops, the stomata close.

Sleeping leaves: potassium content was controlling the spreading and folding of plant leaves in response to changes in day and night, or the rhythm of the day.

Phloem transport: potassium is responsible for helping sucrose enter the phloem and move more solute in the phloem.

For durian planting, potassium helps the fruit's rapid growth, strength, and attractive skin color. It also aids in the resistance of some plant diseases [14].

2.3 Nutrient analyzer

In this study, the nutritional analysis utilized for nitrogen (N) analysis was CNS analyzer, while ICP-OES (inductively coupled plasma optical emission spectrometer) was applied for phosphorus (P) and potassium (K) measurement. These instruments are used

to measure nutrient content in the standard method (AOAC 984.27 for P and K analysis), which is then utilized as a reference value in modeling.

2.3.1 CNS analyzer

CNS analyzer is a quantitative examination of the elements carbon (C), nitrogen (N), and sulfur (S) in a sample. Following sample combustion with oxygen gas to generate an oxidation reaction that provides a mixed gas, the mix gas is separated into a pure gas, the element content of the sample is measured using a Thermal Conductivity Detector, and it is processed with a computer.

2.3.1.1 The principle of CNS analyzer

Combustion with oxygen gas (O_2) is a chemical reaction that occurs when a combustible substance reacts with oxygen, releasing thermal energy. Simultaneously, it is transformed into a compound oxide or complete combustion product, namely carbon dioxide (CO_2) and water (H_2O) [16]. By incinerating samples containing C, N, and S components in the presence of O_2, the following compounds will be formed:

$$2CO + O_2 \dashrightarrow 2CO_2$$

$$N + O_2 \dashrightarrow NO_2$$

$$4NH_3 + 3O_2 \dashrightarrow 2N_2 + 6H_2O$$

$$S + O_2 \dashrightarrow SO_2$$

$$2SO_2 + O_2 \dashrightarrow 2SO_3$$

The combustion technique is used to burn the sample to measure components C, N, and S in CNS analyzer. After weighing the sample into a container and placing it in an auto-sampler, the sample capsule is dropped into a high temperature combustion tube with helium (He) gas as a carrier gas. When passes through O_2, it reacts and releases thermal energy, raising the temperature to 1800°C, allowing for more complete combustion of components C, N, and S. The sample is oxidized to a mixture of gases: N_2, N_xO_y, CO_2, H_2O, SO_2 and SO_3 (Figure 2.13); then, N_xO_y is converted to gas N_2 and SO_3 to gas

SO_2 by a reduction process (Figure 2.14), and H_2O is removed by filtering to separate the gas. The combination gas was converted to pure gas using a GC Column, and the quantity was quantified using a thermal conductivity detector (TCD) as chromatograms of peak C, N, and S that were computer controlled. The resultant chromatography is compared to standard substances, and the results are presented as a percentage (%) of C, N, and S *[17, 18]*.

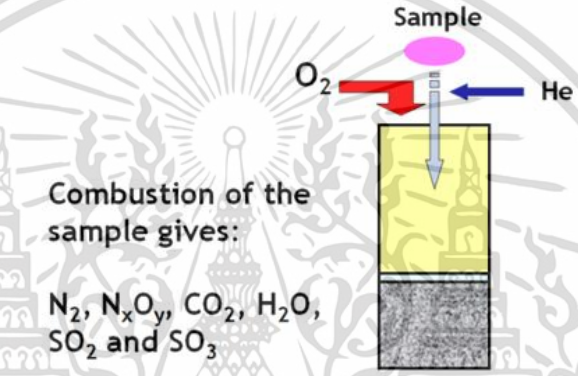


Figure 2.13 Oxidation [18].

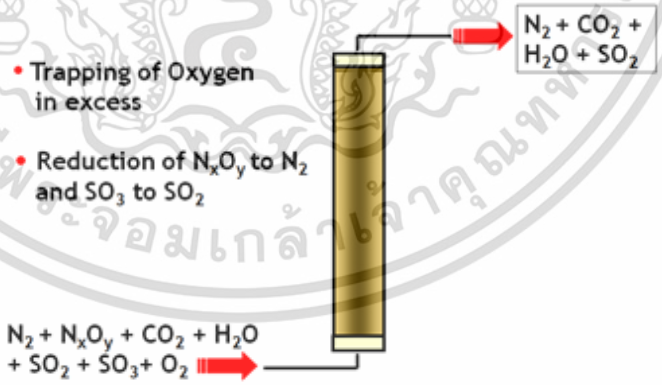


Figure 2.14 Reduction [18].

2.3.1.2 Components of CNS analyzer

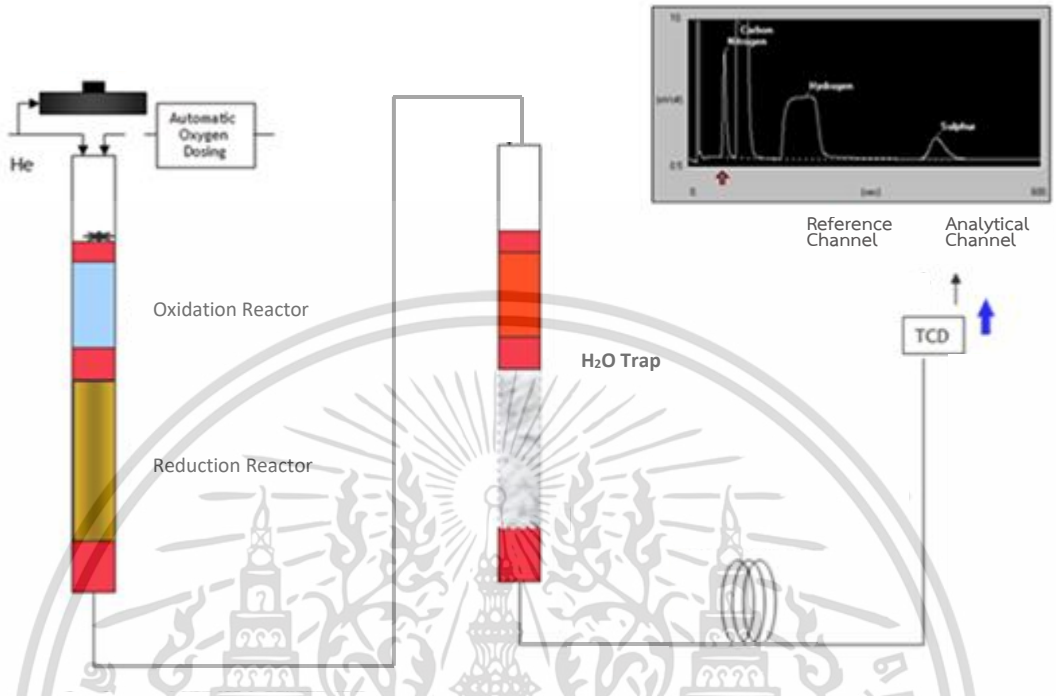


Figure 2.15 Components of CNS analyzer [18].

Components of CNS analyzer (Figure 2.15) have 8 parts [18]:

1 **Auto-sampler** is used to put the sample into the capsule.

2 **Furnace** is used to burn for oxidation.

3 **Oxidation Reactor** using glass wool to reduce heat.

4 **Reduction reactor** using copper and N-catalyst for reduction.

5 **H_2O Trap** using anhydrone to remove water vapor from the mixture gas.

6 **GC Column** serves to separate the mixed gas to pure gas.

7 **Thermal Conductivity Detector (TCD)** measures the amount of pure gas.

8 **Analysis, control, processing and display system**

2.3.1.3 Durian leaves sample preparation for CNS analyzer

To analyze nutrients using a CNS analyzer, the sample must be prepared such that it is small enough to allow for complete combustion. The sample is made in the follows:

1) Dry sample of dried ground leaves at 70°C for 48 hrs.
2) Sieve samples through 0.1 mm.
3) Weigh sample of 0.1000 g and put it in a container.
4) Put a container to CNS analyzer for analyzing.

One sample takes about 3-5 minutes for analysis and 30 minutes for all the process.

2.3.2 ICP-OES (inductively coupled plasma optical emission spectrometer)

The ICP-OES technique is used to evaluate the quantity of phosphorus (P), potassium (K), sulfur (S), calcium (Ca), magnesium (Mg), iron (Fe), manganese (Mn), copper (Cu), zinc (Zn), boron (B), molybdenum (Mo) and nickel (Ni) in a sample by measuring the amount of light energy released by the sample's atoms. When heat energy is applied to atoms, electrons from the ground state become excited state, and when they return to the ground state, they release light energy. The released light energy is then directed to the optical isolator, where the signal is collected, converted into electrical contracts, and the nutritional content is analyzed. The samples used in the analysis must be prepared in the form of a solution and have a limit of measurements at the billion-part scale (parts per billion, ppb).

2.3.2.1 Theory of the released light energy

When providing thermal energy to the atoms that change the state of an electron from the ground state to the excitation state. Atoms are composed of protons, neutrons, and electrons which protons and neutrons combine to nucleus, and electrons orbit the nucleus. The electron arrangement is characterized by orbital. Each orbital has its own unique set of energy levels. The energy levels of orbitals positioned farther from the nucleus are higher. In ground state, the atomic has the lowest energy level and is most stable. Many phenomena occur when energy is applied to an atom. Which ICP-OES technique, electrons absorb energy and change into electrons ground state to the excitation state that called excitation. Electrons from the ground state jump up to another

orbital where higher energy level. Therefore, the atoms are not stable, which the electron attempts to release the energy to back to ground state (Figure 2.16). The energy loss may appear in multiform, 1) energy transferring to other atoms by crashing, 2) energy loss in the form of electric magnets that celled photo, 3) electrons absorbed enough energy to cause the electrons to fall out of the atom, so the atoms are electrically imbalanced, causing them to form a positive charge. Showing excitation ionization and emission, the energy emitted changes the state of radiation in the form of electromagnetic waves or heat (Figure 2.17) [19].

Figure 2.16 Receiving and releasing energy when stimulated atomic [19].

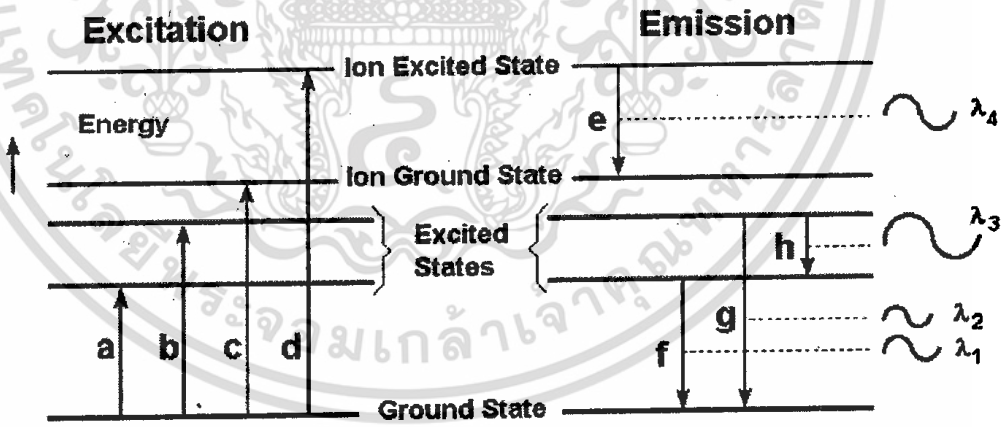


Figure 2.17 Diagrams of energy level when receiving and releasing energy [19].

The difference energy between the high energy level and the low energy level of the state transition indicates the wavelength of the electromagnetic wave related to the transition. It can be calculated using Planck's equations:

$$E = hV \quad (2.1)$$

where E is the energy value (J)

h is the Planck 's constant (6.626 x 10^{-34} J s)

V is the frequency of radiation (s^{-1})

But from the equations

$$V = \frac{C}{\lambda} \quad (2.2)$$

where c is the speed of light (m s^{-1})

λ is the wavelength (m)

therefore

$$E = \frac{hC}{\lambda} \quad (2.3)$$

The energy value is shown to be inversely related to the wavelength. Each element has a unique set of energy levels. As a result, each element has its own set of wavelengths resulting from energy absorption and releasing.

2.3.2.2 The principle of ICP-OES

ICP (inductively coupled plasma) is an energy resource to the atomic emission spectrometer using argon (Ar) gas to generate plasma. Plasma is energy resource for the solution sample, whereby the plasma formed in atmospheric pressure conditions is maintained by receiving energy transferred from electromagnetic fields generated by the induction of radio frequency (RF) electromagnetic waves.

When a high-frequency current is directed to a copper induction coil, and water flows through it for cooling, an electromagnetic field change occurs around the coil. The free electrons in the torch are induced by the electromagnetic field that move in a circle perpendicular to the electromagnetic field. The electrons crash with the atoms of argon gas coming from a quartz glass torch resulting in the cation of argon gas and electrons and

heat that is called plasma. Collisions occur continuously in a chain reaction, thereby allowing the plasma to remain intact. The transmission of energy to plasma is like a transformer, where the induction coil is the primary winding, and the disintegrated gas is like a secondary winding. This characteristic of plasma formation is called inductively coupled plasma discharge (ICP discharge) [19].

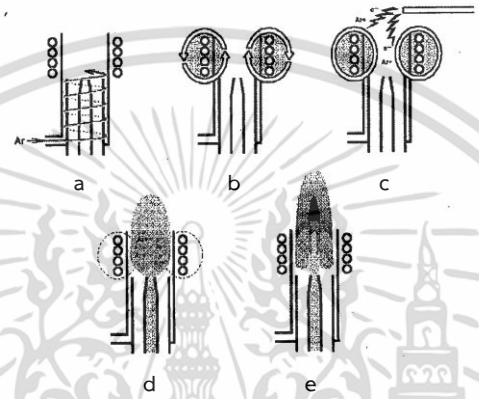


Figure 2.18 ICP torch plasma formation [19].

Figure 2.18 shows step of ICP torch plasma formation [19]:

(a) Releasing argon gas through the torch.

(b) Passing current to the inductor coil.

(c) Changes in the electromagnetic field around the induction coil, causing the electrons in the torch to be induced and move in a circle perpendicular to the electromagnetic field and crash with molecules of argon that causing the disintegration of Ar^{+}+ e^{-}.

(d) Heat production is called plasma.

(e) Injection solution sample through plasma.

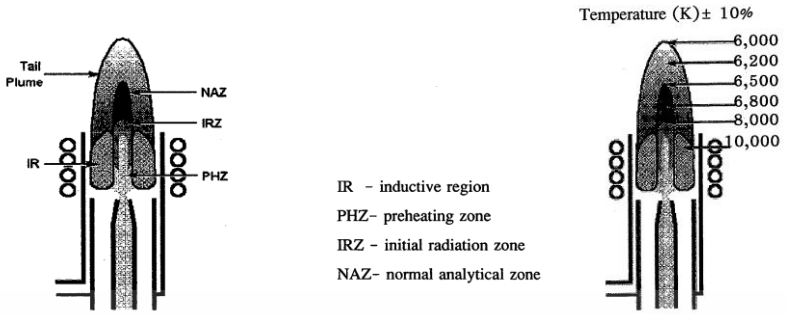


Figure 2.19 Regions of plasma [19].

The regions of the plasma in each zone are called differently (Figure 2.19). The base is toroidal or donut-shaped because the argon gas that flows through and the sample droplets enter the center of the plasma. This region is known as the induction region (IR). Next, the sample droplets are solvent separation at the preheating zone (PHZ), where the plasma separates the solvent from the sample droplets by desolvation, leaving only small solid particles of sample, and molecules are decomposed into gaseous by vaporization, then the molecules are sub-divided into atoms by atomization. The atoms are excited and ionized in the radiation zone (IRZ) by ionization, where temperature is about 8,000°K. The stimulated atoms will emit unique rays. The atoms or ions in the excitation state are atoms or ions with higher energy levels by excitation process, and the normal analytical zone (NAZ) is where the emission line of the element to be analyzed is measured, which has a temperature of 6,800°K [19]. Figure 2.20 shows the process of vaporization, atomization, and ionization of the solution.

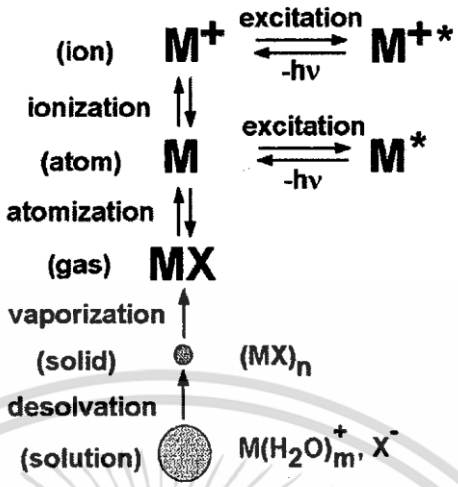


Figure 2.20 The process of vaporization, atomization, and ionization of the solution [19].

2.3.2.3 Components of ICP-OES

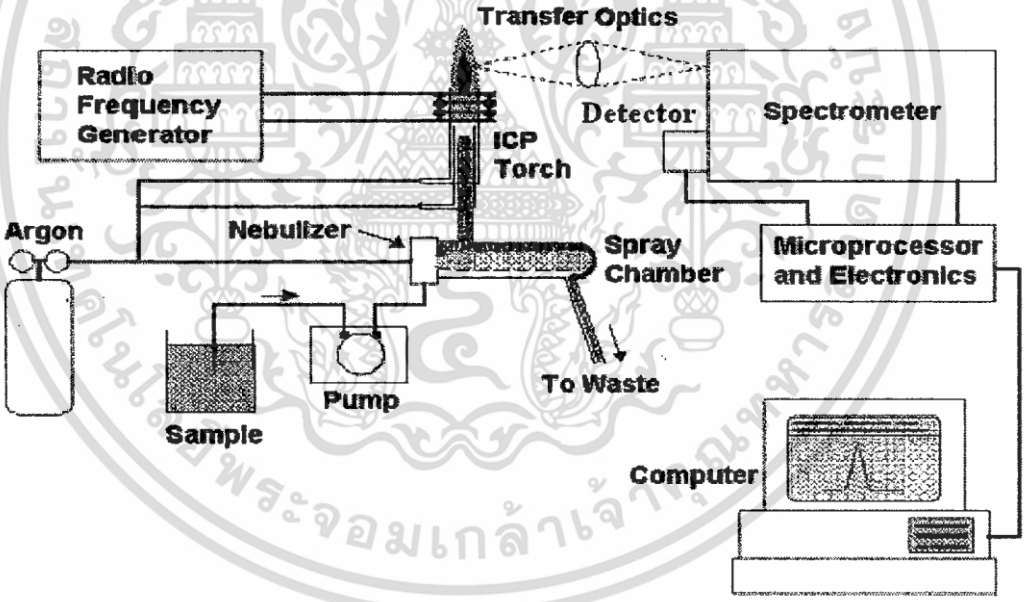


Figure 2.21 Components of ICP-OES [19].

Components of ICP-OES (Figure 2.21) have 5 parts [19]:

1 **A radio frequency generator (RF)** is a device that supplies energy of radio frequency current to the plasma through a coil to produce plasma and keep the plasma in shape.

2 **The CP source** is consists of quartz glass torch, copper inductive coil, and connected to an RF generator that water flows through the torch for cooling. Which is connected to the control system for the pressure and flow rate of argon gas.

3 **The sample introduction** consists of a spray chamber, nebulizer and peristaltic pump that is system for sample introduction into the plasma.

4 **A spectrometer** is a system for separating the wavelength of light that is generated by the light energy emission of atoms or ions of a sample. It is a beam of light with a single wavelength (monochromatic light). This single beam enters the detection to convert the light signal to electric signal, enter it into the signal analysis system and calculate.

5 **System of signal analysis, control process, and display nutrient content.**

2.3.2.4 Durian leaves sample preparation for ICP-OES

To analyze nutrients with the ICP-OES, a mixed solution of a standard solution with a known concentration value (1,000 mg l^{-1}) must be prepared and then analyzed with the ICP-OES for creating a standard graph of each element. The standard graph is used for determining the nutrient concentration in the sample, and the reference solution is used to verify the accuracy of the standard graph in nutrient analysis, when several samples are analyzed.

1 Preparation of standard solution

Table 2.2 Mixture of standard solutions concentrations.

group	nutrient	Standard solution concentration (mg l^{-1})			Volume per 100 ml* (ml)	Initial solution concentration (mg l^{-1})
		high	medium	low		
anion	P	50	5	1	5	1,000
	S	50	5	1	5	1,000
	B	3	0.3	0.06	6	50
	Mo	2	0.2	0.04	4	50
cation	K	200	20	4	20	1,000
	Ca	200	20	4	20	1,000
	Mg	6	6	1.2	6	1,000
	Fe	10	1	0.2	1	1,000
	Mn	10	1	0.2	2	500
	Cu	2	0.2	0.01	4	50
	Zn	2	0.2	0.04	4	50
	Ni	2	0.2	0.04	4	50
Internal standard	Y		2		2	100

*It is a standard solution that is mixed at high concentration.

1) Prepare the solution used to adjust the volume. For leaf nutrient analysis, aqua regia is used as a volumetric modulator, since aqua regia is used to digest leaf samples. Aqua regia was obtained by mixing nitric acid (HNO_3) at a concentration of 37% by weight (14.8 M) and hydrochloric acid (HCl) concentration of 28% by weight (12.0 M) in a ratio of 1 to 3.

2) Prepare the initial standard solution, a known standard material concentration is 1,000 g ml^{-1}, some nutrient need to be diluted to achieve the desired concentration by adjusting the concentration of the element solution as follows:

   - B, Mo, Cu, Zn and Ni, adjust the concentration to 50 g ml^{-1}
   - Mn, adjust the concentration to 500 g ml^{-1}
   - Y, adjust the concentration to 100 g ml^{-1}

3) Prepare a high-concentration standard solution as shown in Table 2.2 by dividing the nutrient into 2 groups: anion group and cation group.

   For the anion group, drop a standard solution of each nutrient in a 100 ml flask is as follows:

   - P, concentration 1,000 g ml^{-1} = 5 ml
   - S, concentration 1,000 g ml^{-1} = 5 ml
   - B, concentration 50 g ml^{-1} = 6 ml
   - Mo, concentration 50 g ml^{-1} = 4 ml

   Firstly, filled Y 100 g ml^{-1} 2 ml into a flask, and volumetrically adjusted with aqua regia for analysis of nutrients at high concentrations, and the other flask is volume-adjusted with aqua regia without Y to be used as a precursor to standard solvent samples at medium concentrations.

   For the cation group, drop a standard solution pipette for each nutrient in a 100 ml flask is as follows:

   - K, concentration 1,000 g ml^{-1} = 20 ml
   - Ca, concentration 1,000 g ml^{-1} = 20 ml
   - Mg, concentration 1,000 g ml^{-1} = 6 ml
   - Fe, concentration 1,000 g ml^{-1} = ml
   - Mn, concentration 500 g ml^{-1} = 2 ml
   - Cu, concentration 50 g ml^{-1} = 4 ml

- Zn, concentration 50 g ml^{-1} = 4 ml
- Ni, concentration 50 g ml^{-1} = 4 ml

Firstly, filled Y 100 g.ml^{-1} 2 ml into a flask, and volumetrically adjusted with aqua regia for analysis of nutrients at high concentrations, and the other flask is volume-adjusted with aqua regia without Y to be used as a precursor to standard solvent samples at medium concentrations.

4) Prepare the mixed standard solution at medium concentration according to Table 2.2 by pipetting 10 ml of high concentration standard solution in a 100 ml flask. filled Y 100 g ml^{-1} 2 ml into a flask, and volumetrically adjusted with aqua regia for analysis of nutrients at medium concentrations, and the other flask is volume-adjusted with aqua regia without Y to be used as a precursor to standard solvent samples at low concentrations by preparing both the base solution of the anion and cation groups.
5) Prepare the standard solution mixed at a low concentration according to Table 2.2 by pipetting 20 ml of medium concentration standard solution in a 100 ml flask. filled Y 100 g ml^{-1} 2 ml into a flask, and volumetrically adjusted with Aqua regia for analysis of nutrients at low concentrations by preparing both the base solution of the anion and cation groups.

2 Preparation of the reference solution

Drop a standard solution at a high concentration of 50 ml in a 100 ml flask, Y concentration 100 g ml^{-1} 2 ml and adjust the volume with aqua regia to be used as a reference solution. The concentration of the reference solution is half the concentration of the standard solution at high concentrations.

3 Preparation of the sample solution

To analyze nutrients with an ICP-OES, the sample must be prepared as a solution as follows:

1) Dry sample of ground leaves at 70°C for 48 hr.

2) Sieve sample through 40 mesh (0.42 mm).
3) Weigh 0.2500 g of sample and put it in the crucible.
4) Burn at 500° C for 6 hr.
5) After burning, drop Aqua regia 10 ml to the crucible and leave it overnight to digest plants.
6) Fill Y concentration (Internal standard) 100 g ml^{-1} 1 ml to the crucible and adjust the volume with Aqua regia to 50 ml.
7) Filter the sample through filter paper no.1.
8) Samples Analysis by ICP-OES.

One sample takes about 4-5 minutes for analysis and 18-20 hrs for all the process.

2.4 Near infrared (NIR) spectroscopy

Near infrared spectroscopy is a technique that uses the principle of interaction, when a near infrared (NIR) wave shines through a material, chemical bonds inside the material absorb energy and cause bond vibrations at each wavelength which is related to the sample's structure.

The NIR technique is a rapid and non-destructive approach. However, it may be employed in variable measurement with no or little sample pretreatment. As a result, the NIR technique is an alternative to the widely used method of verifying the quality of agricultural products.

2.4.1 Principles of Near infrared spectroscopy

Near-infrared spectroscopy (NIRs) is a non-invasive analytical technique used to gather information about the molecular composition and properties of materials. It operates in the near-infrared region at wavelength of 800-2,500 nm or wavenumber 12,500-4,000 cm^{-1} [20]. In NIRs, a light source emits near-infrared light onto a sample, and the transmitted or reflected light is detected and analyzed. The technique relies on the

fact that different molecules absorb and scatter light at specific wavelengths due to their molecular vibrations. The intensity of transmitted or reflected light is measured at specific wavelengths, generating a spectral pattern [21]. The absorption energy assigned to the stretching and overtone/combination vibrations of various bonds such as C-H, N-H, S-H, C-C, C=C, C-N, and O-H in the NIR spectral region [22]. Thus, each sample has a unique spectrum because of different bonds within sample which results of the interaction with NIR radiation are different. Furthermore, particle size and of the samples have an important effect on NIRS [23, 24].

Near-infrared spectroscopy is considered a useful technique in non-destructive and rapid analysis and can be applied to a variety of samples. In addition, NIRs enables continuous monitoring of processes and systems and real-time. Furthermore, it has a possibility to reduce the cost of sample preparation [25]. Therefore, NIRs is nowadays popular for estimating the quality and quantity of samples in food [26], agriculture [27], [28] , etc.

2.4.2 Near-infrared absorbance mode

As NIR radiation passes through or into a sample, the attenuation of the emerging (transmitted or reflected) beam is measured and convert to NIR absorbance [29]. The basic types of NIR absorbance measurement method is transmittance, transflectance, reflectance and interactance [30].

Transmission mode is measured where the light hits the sample on one side and passes it onto the detector [30]. The light passing through the sample is measured on the opposite side (Figure 2.226a). It is suitable for transparent and not thick samples, about 1-50 mm, such as clear liquids contained in quartz tubes, etc.

Transflectance mode is measured where light hits the sample and passes through the sample to a non-absorbing material (ceramic, gold, or aluminum plate) below the

sample, and then reflecting it back to the detector (Figure 2.22b) [30]. It is suitable for aqueous samples or thin materials such as leaf, etc.

Reflectance mode is the light incidents on the surface of the sample propagate in a certain depth first, and then measure the amount of light reflected by detector (Figure 2.22c) [30].

Interactance mode is a cross between transmittance and reflectance. In the case of a fiber optics probe, the light will come out of light ring in the probe to hit the sample, and the light reflected from the sample to the center of the probe (Figure 2.22d) [30]. It is suitable for fruit samples.

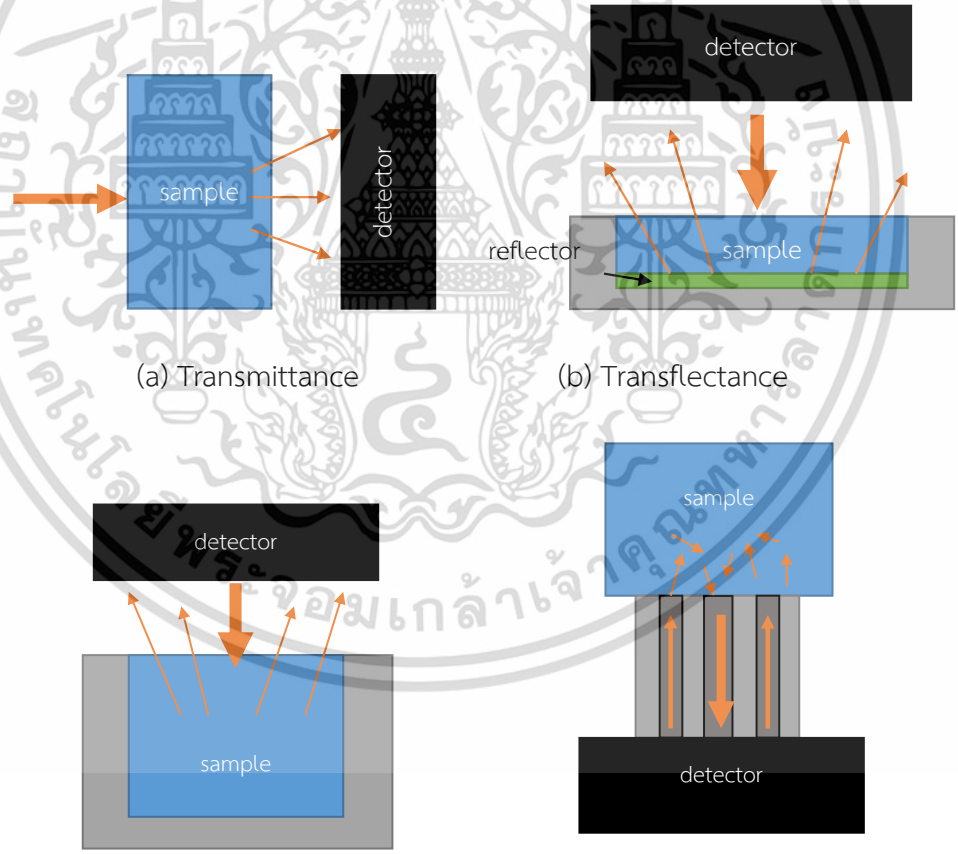


Figure 2.22 NIR absorption measurement methods.

2.4.3 Near-infrared spectroscopy instrument

The principle of the near infrared (NIR) spectrometer is the light passes through a beam splitter to separate the light into different wavelengths and then the light passes through the material. The unabsorbed light is detected by detector and then processed the signal to estimates as a spectrum of absorption. The main components are as follows: light source, beam splitter, detector, and processor (Figure 2.23) [30, 31].

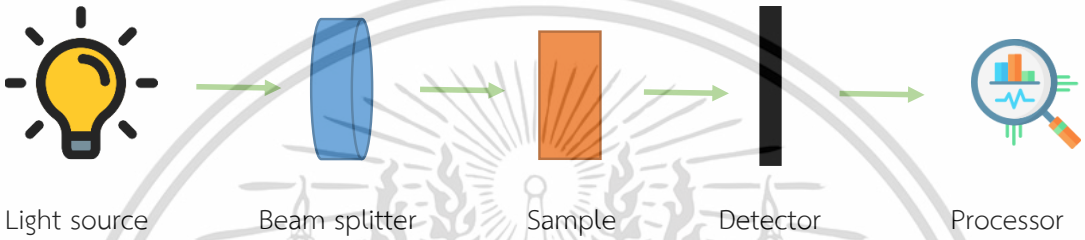


Figure 2.23 Main components of near infrared spectrometer.

Tungsten Halogen lamps are commonly used for light source, which are lamps that rely on heat to the light generation. The lamp contains halogen substances including iodine, chlorine, bromine, and fluorine particles into a glass tube that is made of quartz. This substance prevents sublimation of the filament, which operates at a high temperature of about 3,000-3,400°K, extends the lamp life by 1,500-3,000 hours and has a color index accuracy of 100% [32]. The beam splitter is a device for splitting a beam of light into each wavelength to shin to samples such as optical filter, moving grating, acoustic-optic Tunable Filter (AOTF), Michelson Interferometer, polarization Interferometer or diode arrays [30]. The detector is a device used to measure the light intensity after NIR waves pass through the sample. Commonly used detector materials are PbS (lead sulfide) and InGaAs (lead sulfide) [30]. Processor is to transfer data of the light intensity (after the light passes through the sample) to spectrum for modeling in the future.

In this study, a MPA FT-NIR spectrometer (Bruker Ltd., Germany) was used, which is a near infrared spectrometer with a beam splitting Interferometer [33]. The MPA FT-NIR spectrophotometer has the following advantages: 1) the wavelength covering in the near

infrared region of 12,500-4,000 cm^{-1} (800-2,500 nm). 2) high accuracy due to high signal to noise ratio (s/n). 3) it can be used solid, semisolid, and liquid samples. 4) it can be used in multiple modes including sample compartment, integrating sphere, fiber optic probes, and transmission unit with optional sample wheel. 5) it has an easy-to-use system by the opus program. 6) the detector has a high sensitivity because it uses InGaAs. 7) it has long service life, durable, easy maintenance. and 8) it is suitable for laboratory and system applications to measure the quality of various materials [33].

(a) (b) (c) (d)

Figure 2.24 FT-NIR spectrometer in different modes [33].

Figure 2.24 shows mode of FT-NIR spectrometer [33]:

(a) Sample Compartment: applies with liquid samples that require temperature control.

(b) Integrating Sphere: applies with solid, semisolid or liquid samples. The cup can rotate in case the homogeneous sample.

(c) Fiber Optic Probes: applies with solid or liquid samples located in tanks or requiring direct contact with samples.

(d) Transmission Unit with optional sample wheel: applies with automatic samples.

2.4.4 Near-infrared spectroscopy applications with nutrient in leaf

Near-infrared spectroscopy (NIRs) can be applied to analyze the nutrient content in leaves, providing valuable information about plant health, nutrient status, and physiological processes. There have been several successful studies using NIR to estimate

macronutrients in the leaves of different plants. Table 2.3 summarizes research on NIRs to estimate nutrients in leaves.

Table 2.3 Research on Near-infrared spectroscopy and nutrients in leaves.

Sample	Type	Wavelength (nm)	Analysis	r^2	Reference
N					
Apple	Fresh	350–1100	Rfrog-RF	0.970	Azadnia et al., 2023 [34]
Tea	Fresh	833–2630	VCPA-IRIV-PLS	0.9371	Guo et al., 2023 [35]
Moringa Oleifera	Ground Dry	1000-2500	PLS	0.92	Rébufa et al., 2018 [36]
Olive	Ground Dry	450–1000 1100–1700	PLS	0.73 0.91	Rotbart et al., 2013 [37]
Sugarcane	Ground Dry	-	PLS	0.989	Yarce and Rojas, 2012 [38]
P					
Apple	Fresh	350–1100	Rfrog-RF	0.955	Azadnia et al. 2023 [34]
Rubber	Fresh	350-2500	CRDR-LM-WEVC	0.820	Guo et al., 2022 [39]
Tea	Fresh	900-1750	SPA-MLR	0.9423	Wang et al., 2020 [40]
Sugarcane	Ground Dry	-	PLS	0.988	Yarce and Rojas, 2012 [38]
K					
Apple	Fresh	350–1100	Rfrog-RF	0.956	Azadnia et al. 2023 [34]
Tea	Fresh	900-1750	SPA-MLR	0.9168	Wang et al., 2020 [40]
Moringa Oleifera	Ground Dry	1000-2500	PLS	0.64	Rébufa et al., 2018 [36]
Sugarcane	Ground Dry	-	PLS	0.979	Yarce and Rojas, 2012 [38]

*r^2, coefficient of determination of prediction set; Rfrog, random frog; RF, random forest; PLS, partial least squares; VCPA, core principle of variable combination population analysis; IRIV, iteratively retaining informative variables; LM-WEVC, weighted environmental variables clustering; CRDR, continuum-removed derivative reflectance; SPA, successive projections algorithm; MLR, multiple linear regression

Fresh and ground dry plant leaves were studied for use with the NIR spectrometer. However, it is difficult to determine which sample type is better since both have shown

success. Therefore, we should try using both sample types with durian leaves to make a comparison. Additionally, in the studies conducted by Azadnia et al., 2023 [34], Guo et al., 2023 [35], Guo et al., 2022 [39], and Wang et al., 2020 [40], feature selection wavelengths were used in modeling. The results demonstrate that the performance of the models was either better or similar to the model using the full wavelength. We will incorporate feature selection wavelengths into our model to validate their potential in improving performance.

2.5 Modeling and performance

In the application of the near infrared technique to predict the standard values, modeling is important because it is a tool for prediction. The prediction accuracy depends on the model's performance. Precision models must be based on good data, both spectral data and standard value data. In addition, the modeling algorithm is also an important factor because different algorithms have different procedures and data selections. Therefore, we must choose the appropriate algorithm, and try to apply many algorithms for comparation the model's performance to find the best model.

2.5.1 Steps of modeling

The steps of modeling (Figure 2.25) start with NIR scanning of the samples to obtain spectra, after that the sample was standard analyzed to obtain reference values. Then match spectra and reference value, which this data will be used to model. Next, separating data to two main sets: the calibration set, and the validation set. The calibration set is used for creating a model with algorithm to predict property (classification model) or quality (regression model) of sample. After getting the model, it should be validated to test the model’s performance by validation set. If the test results are reliable, the model can be used to predict.

Caution of modeling

1) The reference values of calibration set must cover in both the minimum and maximum of validation set.
2) The number of samples must be sufficient for modeling that can represent present and future data.
3) The sampling of data sets is important because they are representative of the population that affect the true performance of model.
4) Reference methods should be considered, which should be standardized and accurate methods to obtain accurate reference data because it also affects modeling. The standard error pf laboratory (SEL) must be calculated.
5) The reference values should test the maximum coefficient of determination (R^2_{max}). Which R^2_{max} is highest possible coefficient of determination (R^2) of model, it shows only error of reference method. If R^2_{max} is high that means this reference method can be used, but R^2_{max} is low that means this reference method cannot be used. R^2_{max} is calculated using the following formula:

$$R^2_{max} = \frac{SD_y^2 - SEL^2}{SD_y^2} \quad (2.4)$$

where SD_y is the standard deviation of the calibration set data.

SEL is the repeatability of reference method that can be determined by SD value of the maximum difference in repetition of the same sample for all samples.

6) The NIR scanning should test repeatability and reproducibility. The repeatability of scanning is standard deviation (SD) value of the absorbance of the scanning that determined by scanning the same sample, same position and 10 times, after that selected 3 wavelengths from the overall spectrum and calculated SD average to present the scanning repeatability of spectrometer. The reproducibility of scanning is SD value of the absorbance of the scanning that

is determined by scanning the same sample but re-load sample every time and 10 time, after that selected 3 wavelengths from the overall spectrum and calculated SD average to present the homogeneity of sample. If the sample is homogeneous, it determines the reproducibility of spectrometer.

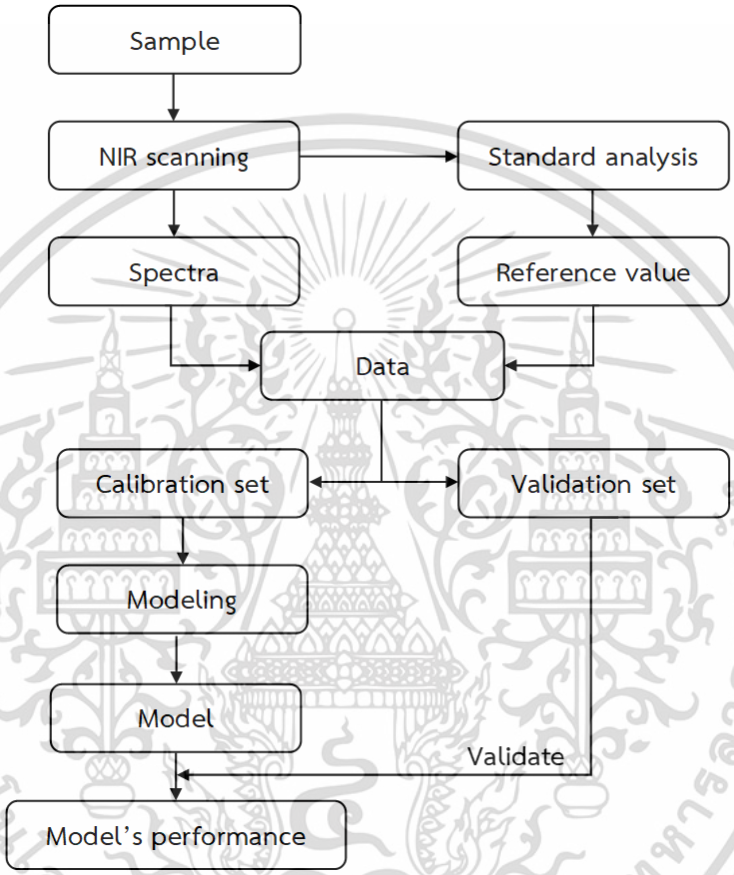


Figure 2.25 Steps of modeling.

2.5.2 Spectra pretreatment

The effect of spectrum includes temperature, humidity, and particle size, which makes spectrum different because of the light scattering. It makes spectrum additive scattering or the multiplicative scattering that affects the spectrum shift all wavelength. In addition, the spectrum obtained overlapping band. Therefore, spectrum treated by mathematically to reduce the discrepancy. Table 2.4 shows MATLAB code for pretreatment methods. Examples of spectra pretreatment include:

1) To remove noise

Savitzky-Golay (S-Golay)

2) To normalize or remove baseline shift

First derivative (1D)

Second derivative (2D)

Standard normal variate (SNV)

Minimum-maximum normalization (MMN)

3) Detrending

Multiple scatter correction (MSC)

Table 2.4 MATLAB code for spectra pretreatment [41].

Pretreatment	Code
Raw	x = raw spectra matrix row is each spectrum column is each wavelength
S-Golay	x_SMT = (sgolay(x', order, framelen))'; where x = raw spectra matrix order = polynomial order framelen = frame length
1D	x_1D = gradient(x)./(- gradient(wave)); where x = raw spectra matrix wave = wavelength
2D	x_2D = (gradient(gradient(x)./(- gradient(wave)))./(- gradient(wave)); where x = raw spectra matrix wave = wavelength

SNV	[n, m] = size(x);
	mean_value = mean(x, 2);
	std_value = std(x, 0, 2);
	temp_mean = mean_value*ones(1, m);
	temp_std = std_value*ones(1, m);
	x_snv = (x-temp_mean)./temp_std;
	where x = raw spectra matrix
	n = number of rows
	m = number of columns

MMN	[n, m] = size(x);
	min_value = min(x, [], 2);
	max_value = max(x, [], 2);
	temp_min = min_value*ones(1, m);
	temp_max = max_value*ones(1, m);
	x_mmn = (x-temp_min)./(temp_max-temp_min);
	where x = raw spectra
	n = number of rows
	m = number of columns

MSC	[n, m] = size(x);
	mean_spectra = mean(x);
	for i = 1:n;
	p = polyfit(mean_spectra, x(i, :), 1);
	x_msc(i, :) = (x(i, :) - p(2)*ones(1, m))./(p(1)*ones(1, m));
	end
	where x = raw spectra
	n = number of rows
	m = number of columns

2.5.3 Calibration model

The calibration model is applied with NIR spectrometer to predict reference value that needs to know from NIR spectra of samples. the prediction values have 2 type that is quantity (number or value) and quality (group, class or location). The Regression model is used for quantity prediction and the classification is used for quality prediction.

2.5.3.1 Regression model

The algorithm for regression model such as Principal Component Analysis (PCA), Principal component regression (PCR), Partial least squares regression (PLS), Support Vector Machines (LSVM) etc. This thesis focused on PLS for regression model (chapter 5).

1 Partial Least Squares Regression (PLS)

The PLS is created from relationship of X matrices (spectra) and Y matrices (reference value) that creates a new variable from a linear combination of the original variables, which is called a latent variable (LV) or component [42]. The number of LVs is less than or equal to the original number of variants. The original X and Y matrices are produced as follows:

$$X = TP^T \quad (2.5)$$

$$Y = TQ^T = XB \quad (2.6)$$

Where P is x-loading ($P^T = (T^TT)^{-1}T^TX$)

Q is y-loading ($Q^T = (T^TT)^{-1}T^TY$)

B is regression coefficient

T is the score that can be obtained from

$$T = XW \quad (2.7)$$

W is x-loading weight

From eq. 2.22 and 2.23 $TQ^T = XB$

$$XWQ^T = XB$$

$$WQ^T = B$$

So,
$$B = W(T^TT)^{-1}T^TY \quad (2.8)$$

Therefore, the predicted valued matrix is calculated as:

$$\hat{Y} = XB = XW(T^TT)^{-1}T^TY \quad (2.9)$$

Table 2.5 shows the MATLAB code for partial least squares regression, which includes calculating the prediction value.

Table 2.5 MATLAB code for modeling [41].

Algorithm	Code
PLS regression	[n, m] = size(x); [Xl, Yl, Xs, Ys, beta, pctVar, mse, stats] = plsregress(x, y, lv, 'cv', n); yfitPLS = [ones(n, 1) x]*beta; where x = raw spectra n = number of rows m = number of columns y = reference value lv = number of latent variables 'cv' = cross-validation method mode Xl = predictor loadings Yl = response loadings Xs = predictor scores Ys = response scores Beta = coefficient estimates for PLS regression pctVar = percentage of variance mse = mean squared error stats = model statistics yfitPLS = predicted value

kNN	Mdl = fitcknn(x, y, 'NumNeighbors', index); flwrClass = predict(Mdl, x); where x = raw spectra y = reference class 'NumNeighbors' = number of nearest neighbors mode Index = number of nearest neighbors Mdl = trained ClassificationKNN classifier flwrClass = predicted class

ANN	net = patternnet(hiddenLayerSize, 'trainscg');
	[net, tr] = train(net, x, y);
	ypred = net(x);
	where hiddenLayerSize = number of hidden layers
	'trainFcn' = scaled conjugate gradient backpropagation mode
	net = pattern recognition network
	x = raw spectra
	y = reference class
	ypred = predicted class

2.5.3.2 Classification model

The algorithm for classification model such as soft independent modeling of class analogies (SIMCA), support vector machines (SVM), k-nearest neighbour (kNN), convolutional neural network (CNN), artificial neural network (ANN), etc. This thesis focused on kNN and ANN for classification model (chapter 4).

1 k-nearest neighbour (kNN)

The kNN algorithm was used to classify the groups. This technique determines the classes that can represent new conditions or cases by checking a certain number. The kNN algorithm of the same or most similar cases or conditions, either the total number of conditions or various cases, will be calculated for each class. New conditions are set to

give the same class as the closest class [43]. Euclidean distances for the calculation is shown below [44]:

$$d\left(x,y_i\right)=\sqrt{\left(x-y_i\right)\left(x-y_i\right)'} \tag{2.10}$$

where $d\left(x,y_i\right)$ is the distance between the sample and nearest neighbour

x is a vector of matric X

y_i is a vector of matric Y

Table 2.5 shows the MATLAB code for k-nearest neighbour, which includes calculating the prediction value.

2 Artificial neural network (ANN)

ANN has the same structural and functional form of processing as the human brain, which adapts itself to the response of the input according to the learning rule. After the network has learned what it needs, it can perform the assigned tasks. The ANN was developed to work like the human brain. It consists of a processing unit called neurone. The number of neurones in the human brain is estimated and connected together [45]. The ANN function is as follows [46];

$$H(x) = \text{bias} + (X_1w_1) + (X_2w_2) + ... + (X_nw_n) \tag{2.11}$$

where H(x) is the sum function of the hidden layer

X_i is the sample i

w_i is the corresponding weights of X_i

n is the number of samples

Table 2.5 shows the MATLAB code for artificial neural network, which includes calculating the prediction value.

2.5.4 Validate method

When the model is obtained, the validate model must be examined to measure the effectiveness of the model to know accurately.

1) **internal validation**

- **Cross-validation**: is a model testing technique that involves leaving one sample out at a time. The model is tested n times, corresponding to the size of the dataset. The process begins by extracting one sample, then creating a model using the remaining examples. Next, the model is tested with the initial extracted sample to evaluate its performance. This process is repeated with different extracted samples until the entire dataset is covered. The final performance measure is typically computed as the average or aggregate of the performance metrics obtained from the n iterations.
- **Test set**: is used for model testing and is created by splitting the data before modeling. The process starts by separating the dataset into a training set and a test set, with a common split ratio being 90-10, 80-20, or as needed. The training set is used to create the model, while the test set is used to test the performance of the model trained on the training set.

2) **external validation**

- **Unknown test**: after internal validation, if the model is available, the unknown test confirms whether the model can accurately predict the next sample. An unknown sample refers to a new sample that is not related to the modeling process, such as a sample collected at a later time or from a different location.

2.5.5 Model's performance

2.5.5.1 Regression model's performance

1 Coefficient of determination (R^2)

R^2 is a value that represents the proportion of variance in an independent variable (X) that can explain the variance in a variable based on (Y). the R^2 of the model can compare with Table 2.6 to determine the function of the model. The R^2 can be obtained from the equation [47]:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \quad (2.12)$$

where R^2 is coefficient of determination

y_i is reference value of each sample

$\hat{y}_i$ is the predicted value of each sample

$\bar{y}$ is the mean of the reference value of the sample

Table 2.6 Guidelines for the interpretation of R^2 [48].

R^2	meaning
<0.66	insufficient for application
0.66–0.81	approximate quantitative prediction
0.81–0.90	good prediction
>0.90	excellent prediction

2 Root mean square error (root mean squares error; RMSE)

RMSEE is root mean squares error of estimation for the calibration set. RMSECV is root mean squares error of cross validation. RMSEP is root mean squares error of prediction. The RMSE can be obtained from the equation [47]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \quad (2.13)$$

where RMSE is model mean square error

3 Standard deviation of error (SE)

SEE is the standard error of estimation for the calibration set. SECV is the standard error of cross validation. SEP is the standard error of prediction. The SE can be obtained from the equation [47]:

$$SE = \sqrt{\frac{\sum_{i=1}^{N} [(y_i - \hat{y}_i) - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)}{N}]^2}{N - 1}} \quad (2.14)$$

Where SE is standard deviation of the prediction error.

4 Ratio of prediction to deviation (RPD)

RPD is the ratio of the standard deviation of the error to the standard deviation. the RPD of the model can compare with Table 2.7 to determine the function of the model The RPD can be obtained from the equation [47]:

$$RPD = \frac{SD}{SEP} \quad (2.15)$$

where RPD is ratio of prediction to deviation

Table 2.7 Guidelines for the interpretation of RPD [48].

RPD	meaning
<2.00	insufficient for application
2.00–2.50	approximate quantitative prediction
2.50-3.00	good prediction
>3.00	excellent prediction

5 Mean error (bias)

Bias is the mean of the error between the reference value and the predicted value from the model. The bias can be obtained from the equation [47]:

$$\text{bias} = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)}{N} \quad (2.16)$$

where bias is mean error with sign

2.5.5.2 Classification model's performance

Figure 2.26 shows a confusion matrix of 3-class classification including class1 = low-level, class2 = appropriate-level, and class3 = high-level. The confusion matrix is an example used to evaluate the classification model. TP is the true positives and E is the misclassification.

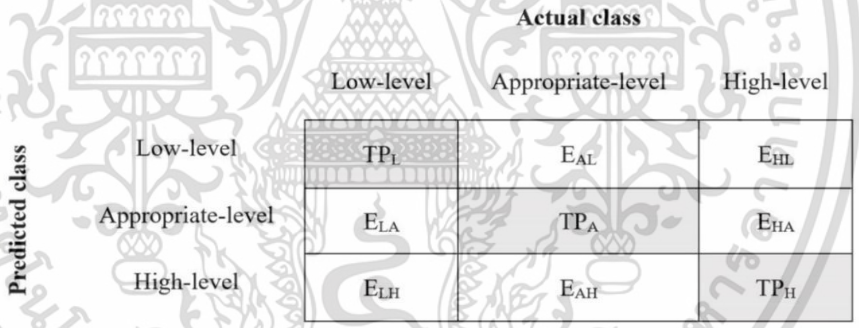


Figure 2.26 Confusion matrix of 3-class classification.

Note

TP_L is the correct prediction of low-level.

TP_A is the correct prediction of appropriate-level.

TP_H is the correct prediction of high-level.

E_{AL} is the actual class being appropriate-level, but the model predicted as low-level.

E_{HL} is the actual class being high-level, but the model predicted as low-level.

E_{LA} is the actual class being low-level, but the model predicted as appropriate-level.

E_{HA} is the actual class being high-level, but the model predicted as appropriate-level.

E_{LH} is the actual class being low-level, but the model predicted as high-level.

E_{AH} is the actual class being appropriate-level, but the model predicted as high-level.

1 Accuracy

Accuracy is the calculation of the number of correct answers compared with the total number of answers given to the model. Accuracy was calculated using the following formula [49]:

$$\text{Accuracy (\%)} = \frac{TP_L+TP_A+TP_H}{\text{Total number of sample}} \quad (2.17)$$

where TP is the true positives

2 Precision

Precision is the ratio of correctly predicted positive observations to the total objectively predicted observations. Precision was calculated using the following formula [49]:

$$\text{Precision}_L = \frac{TP_L}{(E_{AL}+E_{HL})+TP_L} \quad (2.18)$$

$$\text{Precision}_A = \frac{TP_A}{(E_{LA}+E_{HA})+TP_A} \quad (2.19)$$

$$\text{Precision}_H = \frac{TP_H}{(E_{LH}+E_{AH})+TP_H} \quad (2.20)$$

Where Precision_L is the precision of low-level class

Precision_A is the precision of appropriate-level class

Precision_H is the precision of high-level class

E is the misclassification

3 Recall

Recall is the ratio of correctly predicted positive observations to all observations in the actual class or the class accuracy. Recall was calculated using the following formula [49]:

$$Recall_L = \frac{TP_L}{(E_{LA}+E_{LH})+TP_L} \quad (2.21)$$

$$Recall_A = \frac{TP_A}{(E_{AL}+E_{AH})+TP_A} \quad (2.22)$$

$$Recall_H = \frac{TP_H}{(E_{HL}+E_{HA})+TP_H} \quad (2.23)$$

Where $Recall_L$ is the recall of low-level class

$Recall_A$ is the recall of appropriate-level class

$Recall_H$ is the recall of high-level class.

4 F1 score

The F1-score is the average weight of the precision and recall. The F1-score was calculated using the following formula [50]:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.24)$$

2.6 Reference

[1] Office of Agricultural Economics. "Durian Export Statistics" [Online]. Available: https://www.oae.go.th/view. 2023.

[2] Chotwatana, P. **Durian.** Bangkok: Kasetsart University Press. 1987.

[3] Integrated Agricultural Plantation. "The method of artificial pollination Factors Affecting Flowering and Fruiting" [Online]. Available: http://www.kasetkawna.com. 2018.

[4] Suanbanrao. "Soi Dok-Luk Durian Festival" [Online]. Available: http://suanbanrao.com. 2016.

[5] Project to create money. **Increasing productivity, series of durian cultivation.** Bangkok: Utility Publishing Company Limited. 2006.

[6] Chaiwongkiat, D. **Durian '33.** Bangkok, Thailand: Mitsiam Printing House. 1989.

[7] Hiranpradim, H., Chantrapannik, S., and Salakpetch, S. **Durian Production Technology.** Bangkok: Kasetsart University Press. 1998.

[8] Buranachonnabod, B. **Durian orchard.** Bangkok: Agricultural Textbook Production Center for Rural. 2010.

[9] Ministry of Agriculture and Cooperatives, "History and importance of durian," 2015.

[10] Kapook. "How to grow durian trees in pots. Grow fast, yield fast, grow at home" [Online]. Available: https://home.kapook.com. 2015.

[11] Gurukaset. "Durian : flowers burst" [Online]. Available: https://www.gurukaset.com. 2017.

[12] Gurukaset. "Durian" [Online]. Available: https://www.gurukaset.com. 2017.

[13] Buranachonnabod, B. **Durian orchard** Bangkok, Thailand: Agricultural Textbook Production Center for Rural. 2010.

[14] Department of Agriculture. "Nutrient Management and Fertilizing Durian" [Online]. Available: https://www.doa.go.th/share/attachment.php?aid=2975. 2002.

[15] Osotsapar, Y. **Plant nutrition** Bangkok: kasetsart university press. 2015.

[16] Ministry of Energy. **Burning theory.** 2020.

[17] Sitlaothavorn, K. "The use of Argon as Carrier gas in FlashSmart" [Online]. Available: https://www.scispec.co.th/learning/index.php/blog/elemental/argon-carrier-gas-helium. 2019.

[18] Promthong. "CHNS/O analysis" [Online]. Available: http://192.100.77.194/blog/science-equipment/44685. 2017.

[19] Department of Georesources, "Development of methods for analyzing key elemental elements in soil samples by ICP-OES techniques," Bangkok, 2008.

[20] Sandorfy, C., Buchet, R., and Lachenal, G. **Principles of Molecular Vibrations for Near-Infrared Spectroscopy.** 2006.

[21] Leitner, R. and Rosskopf, S. "Identification of Flexographic-printed Newspapers with NIR Spectral Imaging" **World Academy of Science, Engineering and Technology**, vol. 44. 2008.

[22] Foley, W. J., McIlwee, A., Lawler, I., Aragones, L., Woolnough, A. P., and Berding, N. "Ecological applications of near infrared reflectance spectroscopy—a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance" **Oecologia**, vol. 116. 1998. pp.293-305. https://doi.org/10.1007/s004420050591.

[23] Lu, B., Wang, X., Liu, N., Hu, C., Xu, H., Wu, K., Xiong, Z., and Tang, X. "Quantitative NIR spectroscopy determination of coco-peat substrate moisture content: Effect of particle size and non-uniformity" **Infrared Physics & Technology**, vol. 111. 2020. https://doi.org/10.1016/j.infrared.2020.103482.

[24] Minasny, B., McBratney, A. B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., and Joalland, S. "Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon" **Geoderma**, vol. 167-168. 2011. pp.118-124. https://doi.org/10.1016/j.geoderma.2011.09.008.

[25] Wang, X., **book Evaluation Technologies for Food Quality**, 2019.

[26] Oliveira, L. S. and Franca, A. S. **Food Quality: Control, Analysis and Consumer Concerns.** New York, USA: Nova Science Publishers, Inc. 2011.

[27] Vincent, B. and Dardenne, P. **Near-Infrared Spectroscopy.** Springer Singapore. 2021.

[28] Zhang, C. and Su, J. "Application of near infrared spectroscopy to the analysis and fast quality assessment of traditional Chinese medicinal products" **Acta Pharm Sin B**, vol. 4. 2014. pp.182-92. https://doi.org/10.1016/j.apsb.2014.04.001.

[29] Europe, C. d. l. **European Pharmacopeia.** EDQM; 9th edition. 2017.

[30] Terdwongworakul, A. **Non-destructive engineering techniques for agricultural products.** Bangkok: Kasetsart University Press. 2015.

[31] Haruthaithanasan, W. **Near-infrared technology and its industrial applications.** Bangkok: Kasetsart University Press. 2012.

[32] Ngoodat, M., "Light Source," ed, 2018.

[33] Bruker. "Data sheet : MPA II" [Online]. Available: https://www.bruker.com. 2023.

[34] Azadnia, R., Rajabipour, A., Jamshidi, B., and Omid, M. "New approach for rapid estimation of leaf nitrogen, phosphorus, and potassium contents in apple-trees using Vis/NIR spectroscopy based on wavelength selection coupled with machine learning" **Computers and Electronics in Agriculture**, vol. 207. 2023. https://doi.org/10.1016/j.compag.2023.107746.

[35] Guo, J., Huang, H., He, X., Cai, J., Zeng, Z., Ma, C., Lü, E., Shen, Q., and Liu, Y. "Improving the detection accuracy of the nitrogen content of fresh tea leaves by combining FT-NIR with moisture removal method" **Food Chemistry**, vol. 405. 2023. https://doi.org/10.1016/j.foodchem.2022.134905.

[36] Rebufa, C., Pany, I., and Bombarda, I. "NIR spectroscopy for the quality control of Moringa oleifera (Lam.) leaf powders: Prediction of minerals, protein and moisture contents" **Food Chemistry**, vol. 261. 2018. pp.311-321. https://doi.org/10.1016/j.foodchem.2018.04.066.

[37] Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., and Yermiyahu, U. "Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance" **Biosystems Engineering**, vol. 114. 2013. pp.426-434. https://doi.org/10.1016/j.biosystemseng.2012.09.005.

[38] Yarce, C. J. and Rojas, G. "Near infrared spectroscopy for the analysis of macro and micro nutrients in sugarcane leaves" **Sugar Industry**, vol. 2012. pp.707-710. https://doi.org/10.36961/si13611.

[39] Guo, P. T., Zhu, A. X., Cha, Z. Z., Li, M. F., and Luo, W. "A local model based on environmental variables clustering for estimating foliar phosphorus of rubber trees

with vis-NIR spectroscopic data" **Heliyon**, vol. 8. 2022. pp.e09795. https://doi.org/10.1016/j.heliyon.2022.e09795.

[40] Wang, Y.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., and Zhang, Z.-Z. "NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves" **Infrared Physics & Technology**, vol. 108. 2020. https://doi.org/10.1016/j.infrared.2020.103365.

[41] MATLAB. "Help Center" [Online]. Available: https://www.mathworks.com/help/index.html. 2023.

[42] Jiménez-Carvelo, A. M., Martín-Torres, S., Ortega-Gavilán, F., and Camacho, J. "PLS-DA vs sparse PLS-DA in food traceability. A case study: Authentication of avocado samples" **Talanta**, vol. 224. 2021. pp.121904. https://doi.org/10.1016/j.talanta.2020.121904.

[43] Shokrzade, A., Ramezani, M., Akhlaghian Tab, F., and Abdulla Mohammad, M. "A novel extreme learning machine based kNN classification method for dealing with big data" **Expert Systems with Applications**, vol. 183. 2021. pp.115293. https://doi.org/10.1016/j.eswa.2021.115293.

[44] Wang, Y., Pan, Z., and Dong, J. "A new two-layer nearest neighbor selection method for kNN classifier" **Knowledge-Based Systems**, vol. 235. 2022. pp.107604. https://doi.org/10.1016/j.knosys.2021.107604.

[45] Aytaç Korkmaz, S. and Binol, H. "Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection" **Journal of Molecular Structure**, vol. 1156. 2018. pp.255-263. https://doi.org/10.1016/j.molstruc.2017.11.093.

[46] Shaul Hameed, S., Muralidharan, V., and Ane, B. K. "Comparative analysis of fuzzy classifier and ANN with histogram features for defect detection and classification in planetary gearbox" **Applied Soft Computing**, vol. 106. 2021. pp.107306. https://doi.org/10.1016/j.asoc.2021.107306.

[47] Williams, P. **Near-infrared Technology-Getting the best out of light, A Short Course in the Pratical Implementation of Near-infrared Spectroscopy for the User.** Nanaimo Canada: PDK Grain. 2017.

[48] Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., and Mataix-Beneyto, J. "Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils" **Soil Biology and Biochemistry**, vol. 40. 2008. pp.1923-1930. https://doi.org/10.1016/j.soilbio.2008.04.003.

[49] Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Ismail, M. F., Tan, N. P., and Karam, D. S. "Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest" **Computers and Electronics in Agriculture**, vol. 169. 2020. pp.105221. https://doi.org/10.1016/j.compag.2020.105221.

[50] Golodov, V. A. and Maltseva, A. A. "Approach to weld segmentation and defect classification in radiographic images of pipe welds" **NDT & E International**, vol. 127. 2022. pp.102597. https://doi.org/10.1016/j.ndteint.2021.102597.

Chapter 3

Overall precision test for determination the nutrient in durian leaf using near-infrared spectroscopy

Durian planting requires nutrients, particularly nitrogen (N), phosphorus (P), and potassium (K), which are macronutrients. To measure these nutrients, the standard method needs chemicals and takes time for analysis. If near infrared (NIR) spectroscopy can be applied to predict the nutrients concentration in leaf, it can reduce to use chemicals and can quickly predict nutrients concentration with model. The aim of this paper is to test the repeatability and reproducibility for scanning and to determine R^2_{max} for nutrient measurements in durian leaf to check the possibility of modeling and to ensure that this experimental method can be used. From this research, it was found that the repeatability of FT-NIR spectrometer for fresh leaf, dried ground leaves and dried ground leaves pellet was 0.00098, 0.00063 and 0.00036, respectively. The reproducibility of fresh leaf, dried ground leaves and dried ground leaves pellet was 0.01263, 0.00509 and 0.01428, respectively. The R^2_{max} of N, P and K (Table 3.1) were 0.949132, 0.799811 and 0.819135, respectively. The result shows that can used a FT-NIR spectrometer for leaf scanning and a TruMac CNS and an ICP-OES can be used to analyze nutrient concentrations.

*This chapter constituted the publication article: Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. “Overall Precision Test for Determination the Nutrient in Durian Leaf in Durian Orchard using Near-Infrared Spectroscopy” **The 7th Asian NIR Symposium conference (ANS2020)**. at Avani Hotel in Khonkaen, Thailand. February 12-15, 2020.

3.1 Introduction

In 2013-2017, Thailand has increased durian cultivation area from 577,235 rai in 2013 to 592,750 rai in 2017 or 0.73% increase per year. The most popular varieties were Monthong 89.59%, Chani 6.56% and others 3.85% [1]. In durian plantation, farmers must take care closely to get high quality products. Therefore, fertilizing according to soil and plant analysis values will save a lot of fertilizers that must be added to plants, because farmers will apply the nutrients that are insufficient or lacking only. In addition, farmers can also plan for long-term nutrient management.

NIR spectroscopy [2] is a fast and non-destructive sample analysis because no chemicals are used. The spectrum in the near infrared wave will be processed and find the statistical relationship with the data of samples which are analyzed using chemical methods or other methods that has been standardized, and finally get a calibration equation to predict the desired value.

The standard nutrient analyzes of Monthong durian leaf, it requires chemicals and takes time. Therefore, near-infrared spectroscopy techniques can help to reduce chemicals use and can save time to check the nutrient concentration as in the sample scan. It would conserve the environment, reduce the use of resources, help farmers in choosing fertilizer appropriate for the durian tree, reduce production costs and increase profits. Before beginning the experiment, we must do the repeatability, reproducibility and maximum coefficient of determination (R^2_{max}) first, to see the potential trend of the experiment and the precision of the reference measurement.

3.2 Materials and Methods

3.2.1 Sample

The fresh leaf samples were from Monthong durian orchards in Rayong, Chanthaburi and Trad, Thailand. There were 3 samples from a total of 20 leaves per 1

tree per orchard and from 3 different orchards. The samples were washed with HCl 0.1 N acid and distilled water 3 times. After that they were dried at 70 ° C. The durian leaf samples were crushed with a grinder, to have a sieve size of 40 mesh (0.42 mm) to obtain dried ground leaves sample. And 0.15±0.01 g dried ground leaf placed in a stainless block (diameter 1.40 cm) and pressurized at 60 MPa to obtain dried ground leaves pellet.

3.2.2 NIR scanning

The samples were scanned by FT (Fourier transform)-NIR spectrometer (Bruker Ltd., Germany). Each fresh leaf sample was scanned under an aluminum plate at 2 scan positions. The dried ground leaves sample was transferred into a quartz cup of 4.5 cm diameter for full cup and scanned at under cup. The dried ground leaves pellet sample was covered with Spectralon reference material and covered with an aluminum can. For FT-NIR spectrometer, the sample was scanned between wavenumber of 12,500 - 4000 cm^{-1} (800 - 2500 nm) with a resolution of 16 cm^{-1}. The scanning was completed 32 times per one average spectrum. Before each sample scanning, the gold used as a reference material was scanned for background. All scanning was conducted at air conditioning room temperature (25.0 ± 2.0 ℃).

3.2.3 Reference analysis

The samples were analyzed for N by TruMac CNS (Leco, USA) and P, K, Ca, Mg, Fe, Mn, Cu and Zn by ICP-OES (Perkin Elmer, USA).

3.2.4 Repeatability, Reproducibility and Maximum coefficient of determination

The repeatability of NIR scanning was standard deviation (SD) value of the absorbance of time scanning, determined by scanning the same sample, same position and 10 times. We selected 3 wavelengths from the overall spectrum and SD of absorbance in each wavelength to give the repeatability of scanning instrument.

The reproducibility of NIR scanning was SD value of the absorbance of the scanning, determined by scanning the same sample but re-load sample every time and 10 time. We selected wavelength give reproducibility of the scanning like repeatability.

The repeatability of reference method (Rep) is for standard measurement. This can be determined by SD value of the maximum difference in repetition of the same sample for all samples.

The maximum coefficient of determination (R^2_{max}) is the highest possible R^2 for the experiment. This is possible when there is no error in NIR spectrometer. There is only error from the reference test. And R^2_{max} calculated by the following formula [3]:

$$R^2_{max} = \frac{SD_y^2 - Rep^2}{SD_y^2} \quad (3.1)$$

Where SD_y is the standard deviation of the calibration set data.

3.3 Results

Figure 3.1 shows the average spectra of FT-NIR spectrometer, the peak of fresh leaf was selected at 8339, 6896 and 5176 cm^{-1} were CH_3, H_2O and H_2O, respectively, and the peak of dried ground leaves and dried ground leaves pellet were selected at 5793, 5176 and 4003 cm^{-1} were CH_2, CONH and starch, respectively.

Table 3.1 shown repeatability and reproducibility of leaf samples by FT-NIR spectrometer. The repeatability of fresh leaf, dried ground leaves and dried ground leaves pellet was 0.00098, 0.00063 and 0.00036, respectively. The reproducibility of fresh leaf, dried ground leaves and dried ground leaves pellet was 0.01263, 0.00509 and 0.01428, respectively. The R^2_{max} of N, P and K (Table 3.1) were 0.9491, 0.7998 and 0.8191, respectively.

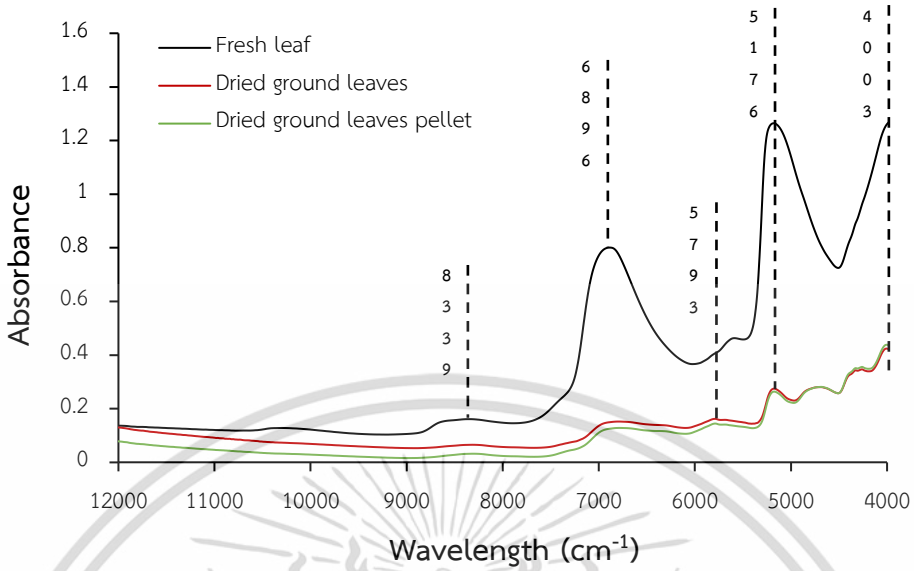


Figure 3.1 The average spectra of fresh leaf, dried ground leaves and dried ground leaves pellet.

Table 3.1 Repeatability and reproducibility of scanning and repeatability of reference method and maximum coefficient of determination (R^2_{max}) for nutrients.

Sample	Scanning		Reference	
	Repeatability	Reproducibility	Nutrient	R^2_{max}
fresh leaf	0.00098	0.01263	N	0.9491
dried ground leaves	0.00063	0.00509	P	0.7998
dried ground leaves pellet	0.00036	0.01428	K	0.8191

3.4 Discussion

Since the FT-NIR spectrometer's repeatability of scanning was low, it could be utilized for fresh leaf, dried ground leaves, and dried ground leaves pellet scanning. This means the FT-NIR spectrometer had high precision in scanning. The repeatability of dried ground leaves pellet was the lowest due to the sample being more or less isotropic homogeneous. Although, the repeatability of fresh leaf and dried ground leaves was higher than the repeatability of dried ground leaves pellet, it was quite low. Therefore, it was

acceptable that the texture is quite uniform. The R^2_{max} of N were quite high, suggesting that a TruMac CNS can be used to analyze N concentrations. However, an ICP-OES can be used to analyze P and K concentrations with caution.

3.5 Conclusion

This research found FT-NIR spectrometer was the highest precision scan dried ground leaves pellet. In reference analysis, a TruMac CNS can be used with high precision for N measurement, but an ICP-OES can be used with caution to analyze P and K concentrations.

3.6 Reference

[1] Office of Agricultural Economics. "History and importance of durian" [Online]. Available: https://www.oae.go.th. 2004.

[2] Gilmer-Osborne, B., Fearn, T., Hindle, P. H., and Hindle, P. T. **Practical Nir Spectroscopy With Applications in Food and Beverage Analysis.** UK: Longman group. 1993.

[3] Saksangium, N. and Sirisomboon, P. "The overall precision test for near infrared scanning and reference method for the determination of soluble solids content and pH of mango for processing factory" **The 13th Thai Society of Agricultural Engineering International Conference (TSAE2020)**. at Kantary Hotel and Serviced Apartment Korat in Nakhonratchasima, Thailand. July 30-31, 2020.

Chapter 4

Rapid measurement of N, P and K concentration levels in durian (CV. Monthong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling

For durian growth to produce high-quality fruit, plants should receive sufficient nutrients. Currently, farmers apply various fertilisers to produce a large quantity and quality of durian fruit, irrespective of the actual nutrients that the plant requires. Accordingly, the production cost is high and non-renewable resources. Therefore, this study focused on rapid classification primary macronutrient levels in durian (Durio zibethinus Murray CV. Monthong) leaves using Fourier transform near-infrared (FT-NIR) spectroscopy and investigated the effect of imbalanced data on efficient classification models. Contents of N, P, and K in durian leaves were measured via NIR with the wavelength range of 800–2,500 nm. Classification models were developed using partial least squares (PLS), k-nearest neighbour (kNN), and artificial neural networks (ANN) with imbalanced and balanced data.

*This chapter constituted the publication article: Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. “Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling” **Measurement.** vol.203. 2022. pp.111975.

The imbalanced data were balanced using a synthetic minority oversampling technique (SMOTE). In this study, the model regarding the fresh leaf sample performed better than that for the dried ground leaf sample. Moreover, the ANN was the best algorithm, exhibiting validation accuracies of classified levels corresponding to N = 0.99 and P = 0.97 when the data were analysed with SMOTE and K = 1.00 from the original balanced data. The imbalanced data affected biased classification when the models could increase the classification accuracy by applying balanced data for modelling.

4.1 Introduction

Durio zibethinus Murray CV. Monthong is a tropical fruit with a unique odour and taste [1]. Southeast Asian exporters in the durian market include Malaysia, Thailand, Cambodia, and Vietnam [2]. Thailand is the largest exporter of fresh and frozen durian whole fruit, freeze-dried durian, and processed products. The export trend of Thai durians appears to increase annually [3]. As there are various competitors, Thai farmers must improve their quality and increase the number of products. Fertilisation, which can help durian plants produce the desired fruits, is important for durian cultivation. Primary macronutrients such as N, P, and K are essential elements usually required for plant growth and maintaining a good overall state [4]. Further, N, P, and K are important for the normal growth and development of plants [5]: N is used for plant growth and development [6]; P affects energy metabolism, nucleic acid synthesis, and photosynthesis [7]; K+ plays a role in protein synthesis, sugar transport, cell growth, and photosynthesis [8]. During durian growth, N directly affects growth, flowering, fruiting, fruit growth, and fruit quality, and aids in the synthesis of proteins and important organic compounds in plants as a constituent of substances responsible for the transmission of energy in processes such as photosynthesis and respiration. K is necessary for fruit trees because it is involved in the synthesis of proteins and carbohydrates in fruits [9]. Orchard fertiliser management

differs at each durian growth stage [10]. If the durian tree lacks N, it will not grow well as there is insufficient food to accumulate for production of flowers and fruits. In contrast, with high-rate N fertilisation, the durian tree either blooms slowly or does not flower [3]. Farmers believe that P fertiliser will activate flowering, but if P is sufficient in the plant, fertilisation only adds additional costs [3]. K deficiency results in smaller fruit size, poor skin colour, and lower acid content and total soluble solids, which hamper the quality of fruit pulp, resulting in tasteless fruit [3]. Because the production cost of durian is high, farmers are more concerned about the quality and quantity of the durian produced; They spray excess fertiliser than the actual amount needed by the plants, which has adverse effects on the growth and quality of durian. Therefore, it is important to control the nutrient content of durian leaves following standard guidelines, which will govern farmers to utilise fertilisers with proper management as per actual requirements. Fertilisation can be adjusted as follows: in cases where leaves have a lower concentration of nutrients, farmers should increase the fertiliser dose by 25%–30% of the normal amount. If the concentration is higher than the standard value, farmers should refrain from fertilisation. If the nutrients are sufficient, farmers can apply fertilisers as usual [9].

Traditional sample preparation for plant nutrient measurement is time-consuming, therefore, farmers are unable to manage fertilisers immediately. Measurement requires the use of chemicals that are harmful to the environment. Near-infrared (NIR) spectroscopy can help solve these problems as it enables fast and accurate determination of nutrient content in plants and non-destructive sample, and reduces the use of chemicals; thus, it is environmentally friendly and has reduced production costs. Fourier transform near-infrared (FT-NIR) spectroscopy is an NIR measurement that uses a fixed spectrometer covered by NIR bands (10,000–4,000 nm [11]), suitable for use in the laboratory. It can be applied to various categories of samples, such as pharmaceuticals, chemicals, petrochemicals, food and beverages, polymers, and agriculture [12]. The NIR spectrum exhibits overtones and combinations of molecular vibrations of–e H, O–H, and N–H bonds

inside the target material [13]. appropriate-levelNumerous studies have successfully applied NIR to estimate the nutrient content in the leaves of several plants [14-26]. NIR can measure N concentration in leaves because compounds in leaves are organic, chlorophyll (C–H bonds), or proteins (C–H and N–H bonds), which exhibit peaks in the NIR spectrum [27]. For P and K contents, NIR cannot directly measure the P or K bonds, but has been successful in predicting the P and K concentrations in leaves. Yu-Jie et al. [28] explained indirect detection of P with NIR from proteins because P is a constituent of nucleic acids, and nucleoproteins are components of proteins. Peaks of proteins and nucleic acids for P content have been used for the prediction of tea leaves. Susan et al. [29] selected the wavelengths of carbohydrates such as sucrose, starch, and cellulose to calibrate the K concentration in grape and orange leaves. Thus, NIR spectroscopy can be used to predict the nutrient levels in durian leaves.

The qualitative model is important for NIR measurement because it involves calculation to predict nutrient levels using NIR spectra obtained via FT-NIR spectroscopy. Qualitative models can be developed using various algorithms, including partial least squares (PLS), k-nearest neighbor (kNN), and artificial neural networks (ANN), which can improve the accuracy of classification models. PLS is a chemometric that can be applied to determine the relationship between data (NIR absorption and nutrient content) [30]. It creates a new variable from a linear combination of both original variables, called a latent variable (LV) or component [31]. Even for a low number of samples, PLS can be applied [32]. Several studies have successfully used NIR data combined with a PLS algorithm for nutrient content prediction in leaves. The kNN is a machine learning algorithm that is simple and widely used for classification [33] because it can classify by distance between the samples (NIR absorption). Specifically, this technique determines classes that can represent new conditions or cases by checking a certain number. The kNN algorithm for the same or most similar cases or conditions was calculated for each class. New conditions were set to provide the same class as the closest class [34]. Only two parameters are

required for calculation: the *k* value and distance [35]. Meanwhile, ANN is a deep learning algorithm, involving supervised learning [36]. It has the same structural and functional form as the human brain, which adapts to the response of the input according to the learning rule. After the network has learned what it needs, it can perform assigned tasks. ANN was developed to work like the human brain. It comprises a processing unit, called a neuron. The number of neurons in the human brain has been estimated and connected [37]. ANN can be applied with non-linear data, which is very helpful because such data are not always linear [38]. Several studies have successfully used NIR data for nutrient content prediction in leaves combined with PLS [15, 19-21], kNN [39] and ANN [40] algorithms. Based on these studies, it appears that the NIR spectrum can be applied to various algorithms for modelling. To examine which algorithms are the most effective in predicting nutrient levels in durian leaves, we compared results from all three methods.

The number of samples from different classes, particularly agricultural produce, is often unequal and asymmetrical. The proportion of samples in each class has a direct effect on classification [41]. This may create problems such as class overlap, small sample sizes, or small disjuncts that damage classifier learning. Therefore, the classifier may ignore minority class samples, and consequently, the classification accuracy may be reduced. Because of the imbalance problem, the model accuracy decreases, and increasing the number of samples from natural sampling might be difficult. Our data had these limitations. Therefore, to improve model accuracy, synthetic samples were added using the synthetic minority oversampling technique (SMOTE) to balance the number of samples. A previous study used SMOTE with NIR spectra to add samples, which could solve the misclassification problems [16, 17, 42-46]. SMOTE was chosen to increase the number of samples because it can help save time and cost for oversampling.

The goals of this study were as follows: 1) to rapidly measure and classify levels of nutrients as low-level, appropriate-level, or high-level in durian leaves using FT-NIR spectroscopy with PLS, kNN, and ANN models; 2) to compare fresh leaves and dried ground

leaf sample modelling; and 3) to determine which type of data (imbalanced or balanced) results in an accurate model for group predictions. There is no model for the classification of nutrients in durian leaves (*Durio zibethinus* Murray CV. Monthong). However, if this measurement is effective, it can help farmers manage the fertilisation of durian trees, which will reduce costs and increase profits. SMOTE is an alternative method for synthesising new samples to reduce the cost of sample collection for modelling.

4.2 Materials and Methods

4.2.1 Study area

This study was conducted in a durian orchard in eastern Thailand. Six orchards of *Durio zibethinus* Murray CV Monthong were considered. Different locations in Monthong were selected for this study: two orchards from Chanthaburi (12° 51' 29" N and 102° 9' 16" E), two orchards from Rayong (12° 49' 60" N and 101° 25' 60" E), and two orchards from Trad (12° 18' 09" N and 102° 30' 45" E). Durian trees older than eight years from each orchard were chosen for the experiment according to Phuwarodom [9]. Thirty durian tree samples were collected. Durian leaf samples collected from December 2018 to November 2020 were used to model the flushing, flowering, blooming, and fruit-setting stages [10] of annual durian production. To check the model performance for real samples, unknown samples from the flushing and flowering stages were collected (December 2020 to February 2021).

4.2.2 Durian leaf sample

In total, 236 fresh durian leaf samples (*Durio zibethinus* Murray CV. Monthong) were collected (one sample in the model included 20 fresh leaves per tree on average; thus, 4,720 leaves were used in total). Fresh durian leaves selected as samples were dark green mature leaves, located 2nd to 3rd from the top of the twig [9]. Fresh leaves were packed in plastic bags and stored in ice buckets with ice at the bottom for shipping to the

laboratory. The samples were protected using paper to avoid direct contamination with the ice.

In total, 236 dried ground leaf samples were collected. Each sample, comprising 20 fresh leaves, was washed thrice with 0.1 N HCl acid and distilled water. Samples were dried at 70°C for 48 h, crushed, and sieved through 40 mesh (0.42 mm) perforated screen [47].

4.2.3 NIR measurement

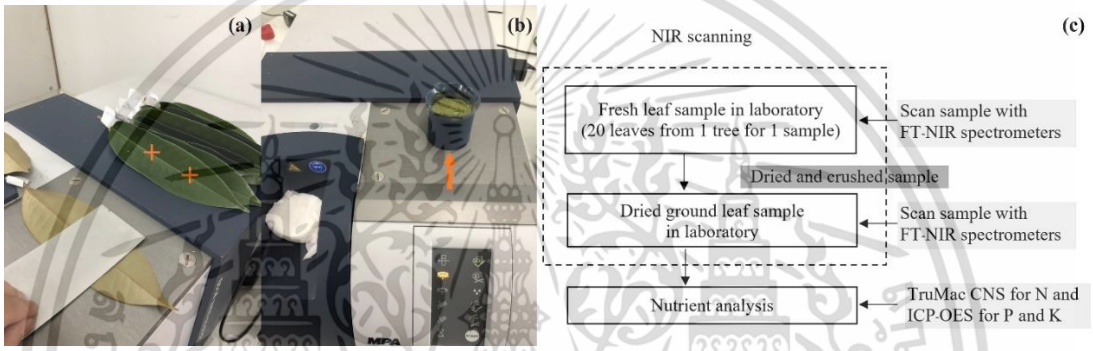


Figure 4.1 Scanning points for fresh leaf (a), dried ground leaves (b), and NIR scanning process (c).

The samples were scanned using an FT-NIR spectrometer in reflectance mode with the wavelength range of 800–2,500 nm (12,500–4,000 cm^{-1}), average of 32 scans, resolution of 6 nm, and gold as the reference material. The measurements were performed under laboratory conditions at 25±2°C. The fresh leaves were scanned at two positions (Figure 4.1a), and the dried ground leaves were scanned under a quartz cup (Figure 4.1b). The durian leaf scanning process is depicted in Figure 4.1c.

4.2.4 Primary macronutrients measurement

The primary macronutrients can only be measured using dried ground leaf samples. To prepare for the measurement, the samples were dried at 70°C for 2 days. The N concentration was measured using a TruMac CNS (Leco, USA) with a 0.1 g dried ground leaf. Meanwhile, the P and K concentrations require further sample preparation. Before

the P and K concentration measurements, each sample was prepared via burning at 500°C for 6 h with a 0.25 g dried ground leaf, digested with 10 ml of Aqua regia, and measured via ICP-OES (Perkin Elmer, USA). The levels of macronutrients were divided by the percent concentration in the leaves, as detailed in Table 4.1

Table 4.1 Macronutrient concentration levels in durian leaf.

Nutrient concentration	Level		
	Low	Appropriate	High
N (%)	< 2.00	2.00 – 2.30	> 2.30
P (%)	< 0.15	0.15 – 0.25	> 0.25
K (%)	< 1.70	1.70 – 2.50	> 2.50

4.2.5 Synthetic minority oversampling technique (SMOTE)

SMOTE provides a new approach to the oversampling of a minority class [48] that will random samples, which finds the k-nearest neighbours of the minority class and creates a synthetic instance [49]. A synthetic instance is the multiplication of random numbers between 0 and 1 with the feature vector of samples [50] to obtain a new synthetic sample. The SMOTE function is expressed as [34]

$$X_{new} = X_i + (\text{rand}(0,1) \times d(X_i - X_m)) \quad (4.1)$$

where X_{new} is the new synthetic sample, rand(0,1) is a random number between 0 and 1, $d(X_i-X_m)$ is the distance between X_i and X_m, X_i is a minority sample, and X_m is the closest sample to sample X_i. In this study, data were balanced using the SMOTE considering the number of nearest neighbours and the oversampling percentage of each datum, as shown in Table 4.2.

Table 4.2 Modelling parameters.

Method	Parameter
SMOTE	Nearest neighbours = 5, oversampling percentage: N; appropriate-level = 200%, low-level = 500%, P; high-level = 150%, low-level = 500%
PLS	Linear regression, LV = 10
kNN	Euclidean distances, k = 1–50
ANN	hidden layers = 10

4.2.6 Modelling

The classification methods used in the study included PLS, kNN, and ANN, and all parameters defined for each model are listed in Table 2. Before modelling, the NIR spectrum data were prepared using different pre-treatment methods, and principal component analysis (PCA) was conducted to decide the most suitable model via comparison. The NIR spectrum data were pre-treated to reduce noise, overlapping peaks, and baseline shifts in the raw spectrum, provided by disturbances with chemical and/or physical (size, temperature, moisture) factors [51] to improve the spectrum. In this study, 13 pre-treatment methods were employed to improve the spectra before modelling, specifically: min-max normalisation, mean normalisation, mean centring, baseline, standard normal variate (SNV), multiplicative scatter correction (MSC), smoothing, 1^{st} derivative, 2^{nd} derivative, smoothing+1^{st} derivative, smoothing+2^{nd} derivative, SNV+1^{st} derivative, and SNV+2^{nd} derivative. PCA reduces real data into a new linearly uncorrelated feature that uses the principle of orthogonal transformation. Accordingly, lower-dimensional data can be obtained while maintaining as much data variability as possible

[52]. PCA was performed on the full wavenumber to reduce the number of variables [53] from the wavenumber of 1,156 to 365 principal component score (PCs).

4.2.7 Performance of the classification model

A 3-class confusion matrix (Figure 4.2) was established to evaluate the classification model. Model comparison was performed based on classification performance, including accuracy and F1 score. Both performances can be calculated by the number of sample predictions in the confusion matrix using the formulae in Table 4.3. The steps of the experiment are summarized in Figure 4.3.

Table 4.3 Formulae for calculating classification performance parameters.

Performance	Formula	Explain
Accuracy (%)	$\frac{\text{Total TP}}{\text{Total number of sample}}$	Calculates the number of correct answers compared with the total number of answers given to the model
Precision (%)	$\frac{TP}{TP+FP}$	Ratio of correctly predicted positive observations to the total objectively predicted observations
Recall (%)	$\frac{TP}{TP+FN}$	Ratio of correctly predicted positive observations to all observations in the actual class or the class accuracy
F1 Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+ Recall}}$	Average weight of the precision and recall

Predicted class \ Actual class	Low-level	Appropriate-level	High-level
Low-level	TP_L	E_{AL}	E_{HL}
Appropriate-level	E_{LA}	TP_A	E_{HA}
High-level	E_{LH}	E_{AH}	TP_H

*TP is true positive of class.
E_{AL}, E_{AL} is false positive of low-level class.
E_{LA}, E_{HA} is false positive of appropriate-level class.
E_{LH}, E_{AH} is false positive of high-level class.
E_{LA}, E_{LH} is false negative of low-level class.
E_{AL}, E_{AH} is false negative of appropriate-level class.
E_{HL}, E_{HA} is false negative of high-level class.

Figure 4.2 Confusion matrix for 3-class macronutrient classification.

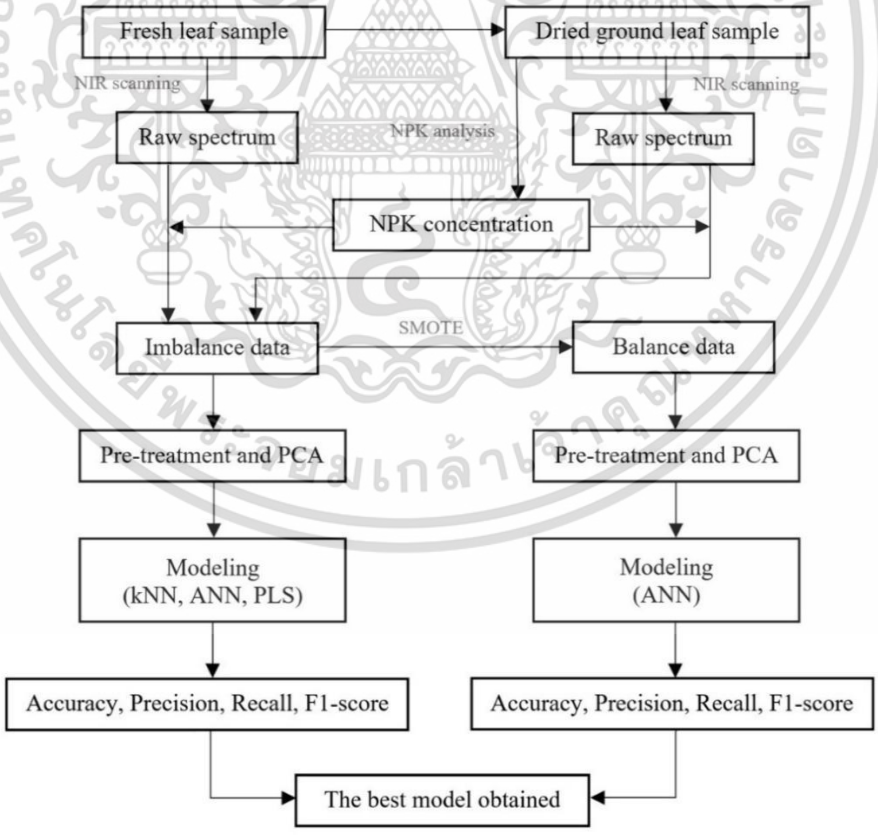


Figure 4.3 Process for obtaining classification model.

4.3 Results

In this study, full wavelengths (12,500–4,000 nm) were used to scan the samples via FT-NIR spectroscopy. The raw spectrum of fresh leaves (Figure 4.4a) exhibits high peaks at 1,440, 1,780, 1,900, and 2,500 nm, which are peaks of starch and cellulose [54]. In the raw spectrum of dried ground leaves (Figure 4.4a), high peaks can be seen at 1,440, 2,323, and 2,500 nm, corresponding to starch and CH_2 [54]. The 1st derivative spectrum of fresh leaves (Figure 4.4b) shows high peaks at 1,395, 1,500, 1,900, 2,310, and 2,461 nm. For the dried ground leaves Figure 4.4b), high peaks are apparent at 1,900, 2,050, 2,310, and 2,461 nm, which are the peaks of CH_2, protein, and starch [54]. The 2nd derivative spectrum of fresh leaves (Figure 4.4c) exhibits high peaks at 1,395, 1,900, 2,280, and 2,323 nm; that of dried ground leaves (Figure 4.4c) shows high peaks at 1,900, 2,280, and 2,323 nm, corresponding to starch, CH_3, and cellulose [54].

Table 4.4 presents the statistics of the original data for N and P contents covered in the three levels (low, appropriate, and high). K covered only two levels (low and appropriate), because the maximum value was 2.35%, which is still at the appropriate-level. Durian leaf samples have low K values because the soil in eastern Thailand is acidic, which causes a loss of K [9, 55]. The average pH of durian tree soil was 4.95.

Table 4.4 Samples statistics.

Nutrient	Statistics			
	Max (%)	Mean (%)	Min (%)	SD (%)
N	4.20	2.48	1.53	0.40
P	0.83	0.23	0.04	0.11
K	2.44	1.64	1.03	0.33

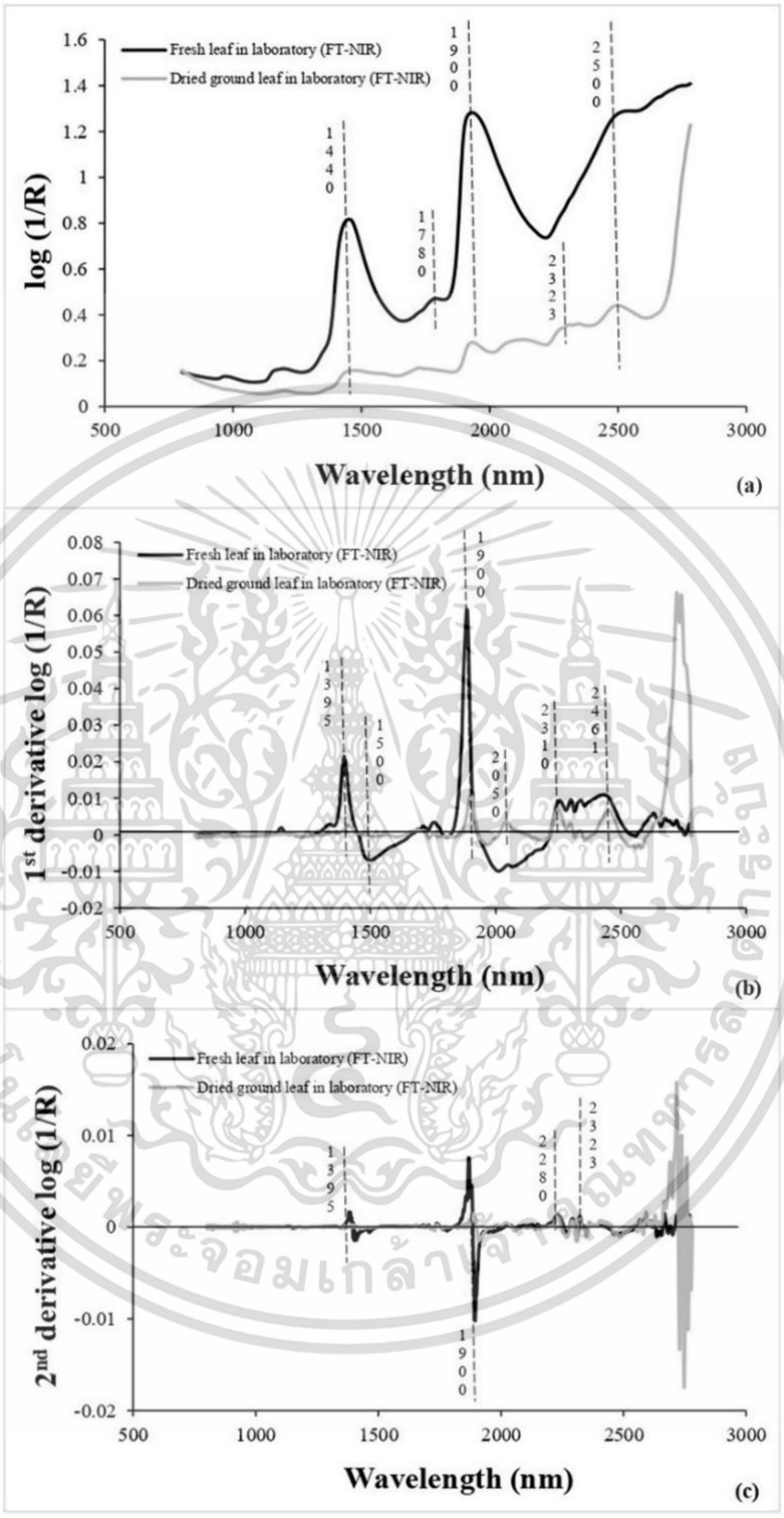


Figure 4.4 Average spectra for raw (a), 1st derivative (b), and 2nd derivative spectra (c).

Table 4.5 Number of samples in each dataset for developing models.

Nutrient	Data (ratio)	Data set	Sample number			
			Low	Appropriate	High	Sum
N	Original	Total set	15	54	167	236
	(2:7:22)	Calibration set	11	37	112	160
		Validation set	2	9	29	40
		Unknown set	2	8	26	36
	SMOTE	Total set	80	147	168	359
	(6:10:10)	Calibration set	58	105	106	269
		Validation set	20	34	36	90
P	Original	Total set	18	156	62	236
	(2:17:7)	Calibration set	14	104	42	160
		Validation set	2	28	10	40
		Unknown set	2	24	10	36
	SMOTE	Total set	98	156	140	358
	(2:3:3)	Calibration set	72	99	97	268
		Validation set	24	33	33	90
K	Original	Total set	136	100	0	236
	(6:4)	Calibration set	93	67	0	160
		Validation set	23	17	0	40
		Unknown set	20	16	0	36

Table 4.5 lists the number of samples in each dataset. The model was validated using an 80% calibration set (160 samples) and 20% validation set (40 samples). Thirty-six unknown samples were considered for external validation. The N class was in the ratio 2:7:22 for low-, appropriate-, and high-level classes, respectively. Most of the collected

samples had high N concentrations and the data were imbalanced. For P-levels, 2:17:7 is the ratio of the low-, appropriate-, and high-level classes, respectively. P-values from most samples were at the appropriate-level, and the data were imbalanced. However, the K-class had only two levels: a 6:4 ratio of low-level to appropriate-level, whose data were approximately balanced. Nutrient distribution trends in the calibration, validation, and unknown sets were similar.

4.3.1 Evaluation of models with original data (imbalance class)

Table 4.6 lists the best macronutrient classification accuracies of the durian leaf models using the original data. All nutrients were modelled using kNN, ANN, and PLS algorithms. In classifying the N level, the ANN method seemed the most accurate. All models exhibited an accuracy of validation greater than 0.85, with the highest of 0.93 obtained from the spectra treated by smoothing+1st derivative without PCA. When testing the selected model with an unknown sample set, the accuracy of 0.94 was achieved.

A model was developed with the highest accuracy of 0.93 to predict the P concentration level. Dried ground leaf samples and min–max normalization spectra with PCA and ANN algorithms were adopted in the model. Because dried samples must be used in this model, grinding the samples requires time. The model using fresh leaf was slightly less accurate than that using the dried ground leaf, with the accuracy of 0.88 for validation. Therefore, the fresh leaf model was chosen with 1st derivative spectrum pre-treatment without PCA to predict P levels because fresh leaf samples were easy to use for prediction. When testing this model with an unknown sample set, the model accuracy of 0.86 was obtained. The classification of the K concentration group was the same as that for N and P. As a result, the ANN algorithm exhibited the highest performance. The best model, created using fresh leaves and SNV+2nd derivative spectrum without PCA, had a validation accuracy of 1.00. When testing this model with an unknown sample set, the

Table 4.6 Model accuracy for classifying levels of macronutrient concentrations in durian leaf.

Nutrient	Sample	SMOTE	Algorithm	Pre-treatment	Accuracy		
					Cal	Val	Unk
N	FL	-	PLS	SNV	0.74	0.85	0.81
		-	kNN	MC	0.71	0.75	0.81
		-	ANN	SNV+1D	0.90	0.93	0.94
		+	**ANN**	**SNV+2D**	**0.97**	**0.99**	**0.97**
	DL	-	PLS	MN	0.83	0.90	0.86
		-	kNN	MMN	0.71	0.73	0.78
		-	ANN	SMT+2D	0.86	0.93	0.81
		+	**ANN**	**SNV+1D**	**0.97**	**0.97**	**0.94**
P	FL	-	PLS	MSC	0.68	0.75	0.61
		-	kNN	SMT+2D	0.75	0.75	0.75
		-	ANN	1D	0.84	0.88	0.86
		+	**ANN**	**SMT+2D with PCA**	**0.94**	**0.97**	**0.97**
	DL	-	PLS	SMT+1D	0.76	0.63	0.69
		-	kNN	RS	0.71	0.78	0.75
		-	ANN	MMN with PCA	0.93	0.93	0.86
		+	**ANN**	**2D**	**0.96**	**0.99**	**0.97**
K	FL	-	PLS	2D with PCA	0.88	0.80	0.76
		-	kNN	SMT+2D	0.71	0.78	0.88
		-	**ANN**	**SNV+2D**	**0.93**	**1.00**	**0.89**
	DL	-	PLS	MMN	0.89	0.75	0.79
		-	kNN	1D	0.74	0.70	0.71
		-	**ANN**	**SNV+2D**	**0.85**	**0.90**	**0.83**

*Cal = calibration; Val = validation; Unk = unknown; FL = fresh leaf scanned; DL = dried ground leaf scanned; kNN = k-nearest neighbour; ANN = artificial neural network; PLS = partial least squares; RS = raw spectrum; 1D = 1st derivative; 2D = 2nd derivative; MC = mean centring; MMN = min-max normalisation; MN = mean normalisation; MSC = multiplicative scatter correction; SMT = smoothing; SNV = standard normal variate; PCA = principal component analysis.

model accuracy of 0.89 was obtained. The ANN models exhibited the highest accuracies because they described both linear and nonlinear relationships [56].

Figure 4.5 displays a scatter plot between PC1 of the original data and the macronutrient concentration, in which the data were nonlinear. Therefore, the ANN algorithm is suitable for these data.

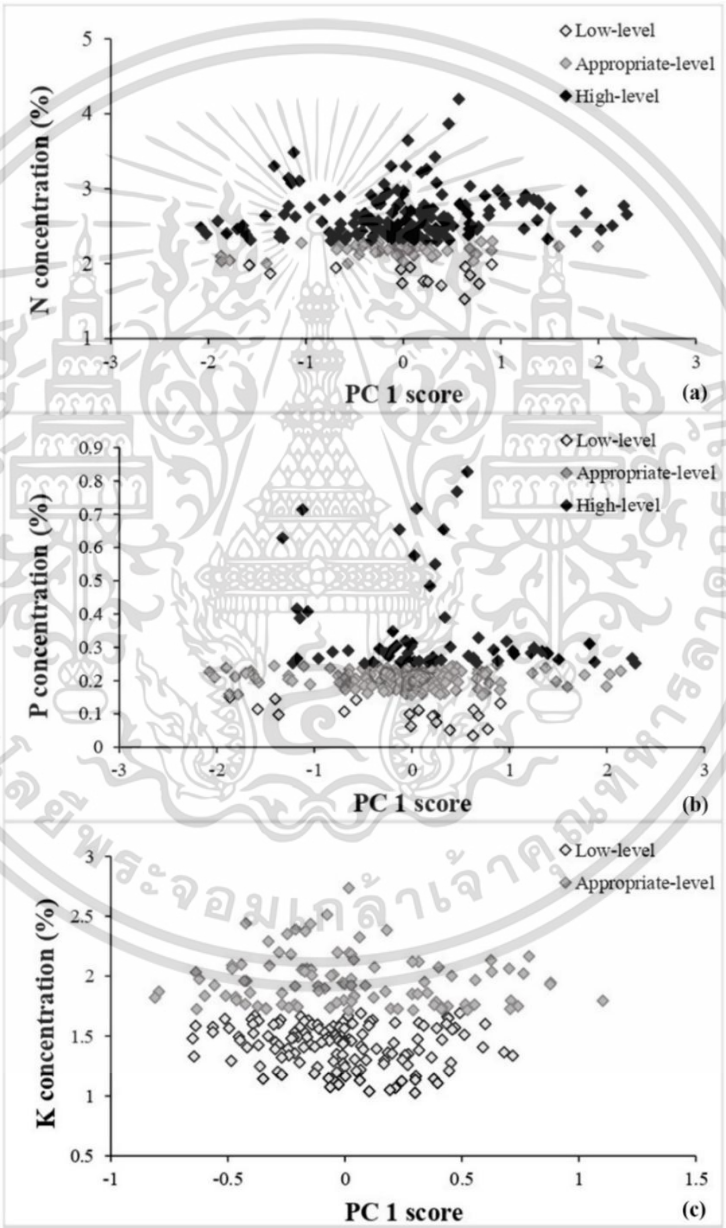


Figure 4.5 Scatter plots between PC1 of original spectral data and macronutrient concentrations.

Table 4.7 ANN model performances for classifying levels of macronutrient concentrations in fresh durian leaf.

Nutrient	Class	F1-score					
		Original			SMOTE		
		Cal	Val	Unk	Cal	Val	Unk
N	Low	0.84	0.67	0.67	1.00	1.00	1.00
	Appropriate	0.85	0.89	0.88	0.97	0.99	0.93
	High	0.94	0.95	0.98	0.97	0.99	0.98
P	Low	0.78	1.00	0.67	0.98	0.98	0.80
	Appropriate	0.93	0.91	0.90	0.92	0.92	0.98
	High	0.74	0.76	0.71	0.94	0.94	1.00
K	Low	0.94	1.00	0.95	-	-	-
	Appropriate	0.92	1.00	0.93	-	-	-

*Cal = calibration; Val = validation; Unk = unknown.

Table 4.7 presents the classification performances of the best ANN models with the original fresh durian leaf data. For the classified N level model, the accuracy was 0.93. F1 scores for high-level samples were greatest in every data set, indicating that the selected model would predict high-level samples with the highest accuracy. In predicting the appropriate-level samples, F1 scores were more than 0.85 for every data set. This demonstrates that the developed model can predict appropriate-levels to a good extent. However, for low-level samples, the prediction result was poor because of the small number of samples considered. The results of the classification of the P-levels, which had the validation of 0.88, showed good F1 scores. Thus, this model can be used to accurately predict levels greater than 0.9. However, predictions of low and high levels were adequate because the number of samples was low. In predicting the concentration of K, besides

the high accuracy of 1.00, the precision, recall, and F1 score of the model were greater than 0.92. This indicates that the model can predict K levels accurately and precisely.

4.3.2 Evaluation of models with SMOTE data (balance class)

After the data were evaluated via SMOTE, the number of samples in the minority class increased (Table 4.4). The SMOTE was performed with the spectrum of the calibration and validation sets of the original data for the N and P samples (imbalance class). The K samples were balanced; therefore, SMOTE analysis was not performed. The ratio of the number of samples in the N concentration was more balance from 2:7:22 to 6:10:10 with a proportion of low-level: appropriate-level: high-level. Appropriate- and high-level samples were similar, but the number of samples at low levels was still lower than those of the others. Owing to the limitations of data generation using SMOTE, only 500% of the original data were generated [48]. The original data of the P concentration model had the low-level–appropriate level–high-level ratio of 2:17:7. Following SMOTE analysis, the ratio became 2:3:3, and the number of samples of the appropriate and high levels was similar. However, the number of low-level samples was still lower than that of the other samples. The maximum and minimum values of the SMOTE dataset for both N and P remained the same as those of the original data. In particular, the ANN method was selected for the analysis because it yielded the highest grouping accuracy results (Table 4.6). The best model from the SMOTE data to predict N concentration levels was generated from fresh leaf samples using an SNV+2^{nd} derivative spectrum, with the accuracy of 0.99 for the validation set. Compared with the best model of the original data (0.93), the accuracy increased by 6%, and the data were more balanced. When testing this model with an unknown sample set, that is, when an unknown sample spectrum was inputted into the model, the accuracy of 0.97 was obtained, equivalent to a 4% increase. The best model for predicting the P concentration level was generated from dried ground leaf samples with the 2^{nd} derivative spectrum; the accuracy was 0.99 for the validation set. When

comparing the accuracy values of the best model from the original data (0.93), the accuracy was found to increase by 6%, and the data were more balanced. The model created using the smoothing+2nd derivative spectrum with PCA showed no significant difference in performance, with the accuracy of 0.97 for the validation set. When testing this model with an unknown sample set, an accuracy of 0.97 was obtained, an increase of 11%.

classifying macronutrient concentrations in fresh durian leaves. For N, the classification accuracies for low-level increased from 0.84 to 1.00 using the SMOTE data instead of the original data (16% increase), appropriate-level increased from 0.85 to 0.97 (12% increase), and high-level increased from 0.94 to 0.97 (3% increase). Similarly, when testing the model with the validation dataset, the accuracies for predicting low, appropriate, and high levels increased by 33%, 10%, and 4%, respectively. When testing the model with an unknown set without SMOTE, it was found that the classification accuracy increased by 33% at a low-level and 5% at an appropriate-level. However, the accuracy of 98% remained at a high level. For P, the classification accuracy increased from 0.78 to 0.98 for the low-level group (20% increase), and from 0.74 to 0.94 (20% increase) for the high-level group. However, the accuracy decreased from 0.93 to 0.92 (1% decrease) for appropriate-level samples. When the model was tested with the validation dataset, it was found that the prediction accuracy increased in the appropriate- (4%) and high-level (20%) groups. However, accuracy remained low at 100%. When testing the model with unknown sample sets without SMOTE, the accuracies of predicting low-, appropriate-, and high-level samples increased by 13%, 8%, and 29%, respectively.

Figure 4.6 depicts the confusion matrix of the best model for classifying the levels of fresh durian leaves. When comparing the confusion matrix between the original data model and SMOTE data model, the original data model (Figure 4.6a) of the N-level with the validation set showed the misclassification of only 2.6%. Additionally, the appropriate level was predicted to be a high-level group of one sample, low-level to be a high-level group of

one sample, and high-level to be an appropriate-level group of one sample. Thus, this model predicts the bias for the majority class (high-level group). However, the model using SMOTE data (Figure 4.6d) solved this misclassification. Moreover, the model using original data (Figure 4.6b) misclassified the P-level; The appropriate level was predicted to be a high-level group of three samples and an appropriate-level group of two samples, adding complexity between the appropriate- and high-level groups. However, the SMOTE data model (Figure 4.6e) solved this misclassification.

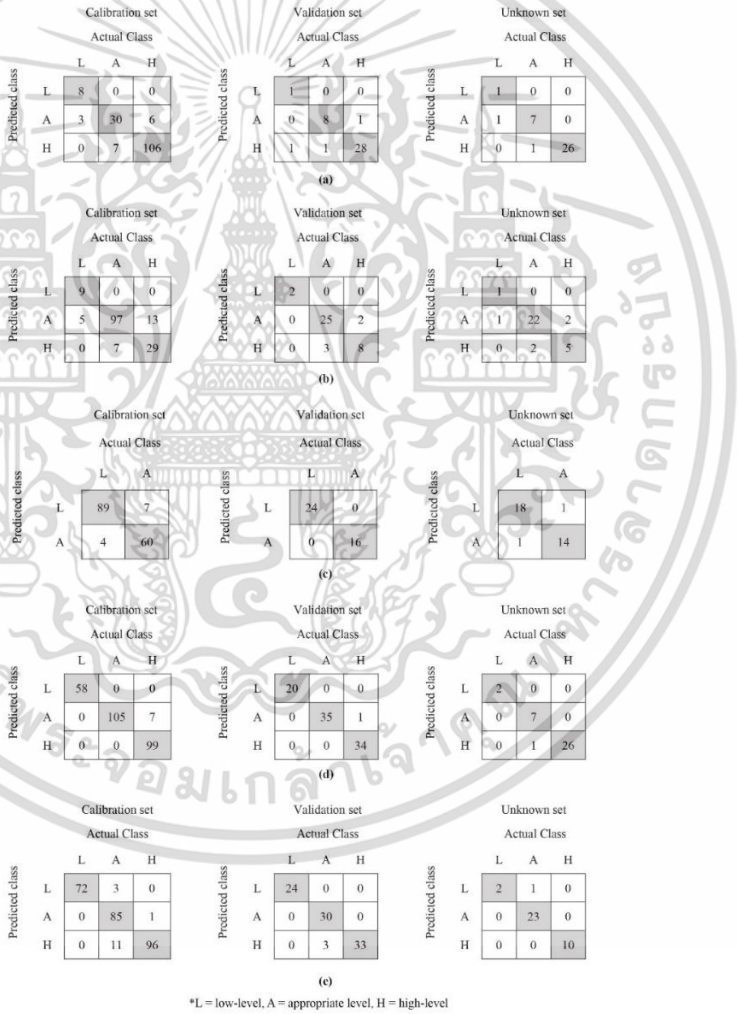


Figure 4.6 Confusion matrixes of the best model for classifying levels of N (a), P (b), and K (c) concentrations with original data, and N (d) and P (e) concentrations with SMOTE.

4.4 Discussion

The absorption bands of durian leaves are related to the leaf components. N is related to chlorophyll and proteins [57], P is related to proteins [28] and chlorophyll [14], and K is related to carbohydrates such as sucrose, starch, and cellulose [29]. NIR can detect chlorophyll (CH_2 and CH_3 [58]), sucrose, starch, proteins, and cellulose (proteins and cellulose are cell wall components [59]). FT-NIR spectroscopy can relate the absorption bands with respect to N, P, and K, and thus, can be used to measure the corresponding contents in durian leaves.

The above result demonstrates that the P classification of fresh durian leaves was more accurate when PCA was used with the spectrum. Because P cannot be directly detected via NIR absorption, it is related to proteins [28]; the protein peaks in the durian leaf spectrum were small and unclear. Thus, the spectrum was transformed into a PC that reduced the number of variables orthogonal to each other (no multicollinearity, so there were no correlated independent variables [60]). The informative data of the sample spectrum were primarily included in the PC score of the preceding PC variable; whereas, succeeding studies included more noise or uninformative data. Variables without multicollinearity and containing information help to make better correlations with protein concentrations in durian leaves. Unlike N and K, which exhibited clear peaks in the NIR absorption of chlorophyll and carbohydrates, PCA transformation did not result in a better correlation.

As a result, the validation accuracy of fresh leaf measurement was slightly higher than that of dried ground leaves regarding N. This is because, as the temperature increases, the chlorophyll degradation rate increases [61] and N is related to chlorophyll [38]. For the P-model, the validation accuracy of the dried ground leaf measurements was slightly higher than that of the fresh leaf samples. Leaf proteins contain both insoluble and soluble proteins in equal amounts [62] and the removal of water makes the protein peak

stand out (at 2050 cm^{-1} (Figure 4.4b)) because the P concentration is related to protein [15]. The K concentration is related to starch [16], in which fresh leaf samples absorb more NIR radiation in the starch band (1900 cm^{-1} (Figure 4.4a, b, c)) than the dried ground leaf sample, which is more affected by the model. Therefore, the fresh leaf model was more accurate than the dried ground leaf model in measuring K concentrations.

The chosen models with original data for predicting nutrient concentrations were created using fresh leaf spectra and ANN algorithm. The number of samples in each class was imbalanced, and each group affected the precision of the model [63, 64]. Due to natural sampling in the durian orchard samples, the number of K-classes was more balanced compared with the N- and P-classes. As a result, the classification precision of K was higher than that of N and P. Further, the model achieved higher values of other performance parameters, including the precision, recall, and F1 score of each class, ensuring the reliability of the model [65]. From the experiments, it was found that the balanced dataset was more accurate than the unbalanced dataset. Amiratul et al. [16] considered balance accuracies of below 40.00%, 40.00%–80.00%, and above 80.00% as poor, moderate, and robust, respectively. When these balanced accuracies were compared with those of the best models in this study, it was found that the balanced accuracies of N, P, and K were in the range of 0.96–1.00, hence these models are robust.

Therefore, the number of samples in each class should be similar to reduce the classification problem caused by the group bias [66]. It can be concluded that using synthetic samples to balance the number of samples in each group can improve modelling only for naturally biased or skewed distribution data. The clustering accuracy of the proposed model was greater regarding classification of the N and P concentrations in durian leaves than when using imbalanced data, which corresponds with YongBae's research [67].

In future, the addition of real samples is important for improving the concentration range covered by the model; in particular, samples with high K concentrations. Moreover,

we will define the treatment of nutrient concentration levels using foliar fertiliser to prepare samples with high concentrations. However, for low-concentration-level samples that are difficult to mark, the SMOTE is a good choice and a proven alternative to real samples by unknown validation.

4.5 Conclusion

This study presented the possibility of using NIR measurements to classify the concentration levels of N, P, and K in durian (*Durio zibethinus* Murray CV. Monthong) leaves. Based on this research, 1) the developed macronutrient-level classification model can be applied for rapid measurement in durian leaves using an FT-NIR spectrometer. The model developed using the ANN algorithm resulted in high accuracy and good precision; 2) The model incorporating fresh leaves exhibited higher accuracy than that for the dried ground leaf because the fresh leaf spectra contained peaks related to N, P, and K; 3) Use of balanced data can obtain higher F1 score and accuracy than use of the original imbalanced data. Analysing the data using the SMOTE, the data were more balanced, and the problem of minorities was addressed without addition of real samples, which could reduce the cost of sample collection. The FT-NIR spectrometer and ANN models with balanced data were accurate for rapidly measuring macronutrient levels in durian leaves. However, the range of K concentration used in this experiment was not sufficient. In future, the macronutrient concentration range of samples should be regarded for a wide range of applications. Moreover, wavelength selection is necessary to improve the model because it only considers nutrient wavelength information.

4.6 Reference

[1] Saechua, W., Sharma, S., Nakawajana, N., Leepaitoon, K., Chunsri, R., Posom, J., Roeksukrungrueang, C., Siritechavong, T., Phanomsophon, T., Sirisomboon, P., Lapcharoensuk, R., and Pornchaloempong, P. "Integrating Vis-SWNIR spectrometer in a conveyor system for in-line measurement of dry matter content and soluble solids content of durian pulp" **Postharvest Biology and Technology**, vol. 181. 2021. pp.111640. https://doi.org/10.1016/j.postharvbio.2021.111640.

[2] Zhou, X., Wu, H., Pan, J., Chen, H., Jin, B., Yan, Z., Xie, L., and Rogers, K. M. "Geographical traceability of south-east Asian durian: A chemometric study using stable isotopes and elemental compositions" **Journal of Food Composition and Analysis**, vol. 101. 2021. pp.103940. https://doi.org/10.1016/j.jfca.2021.103940.

[3] Office of Agricultural Economics. "Durian Export Statistics" [Online]. Available: https://www.oae.go.th/view. 2023.

[4] Maathuis, F. J. **Plant Mineral Nutrients Methods and Protocols.** Humana Press: 2012.

[5] Silva, T. C., Bertolucci, S. K. V., Carvalho, A. A., Tostes, W. N., Alvarenga, I. C. A., Pacheco, F. V., de Assis, R. M. A., Honorato, A. d. C., and Pinto, J. E. B. P. "Macroelement omission in hydroponic systems changes plant growth and chemical composition of Melissa officinalis L. essential oil" **Journal of Applied Research on Medicinal and Aromatic Plants**, vol. 24. 2021. pp.100297. https://doi.org/10.1016/j.jarmap.2021.100297.

[6] Ma, C., Ban, T., Yu, H., Li, Q., Li, X., Jiang, W., and Xie, J. "Urea Addition Promotes the Metabolism and Utilization of Nitrogen in Cucumber" **Agronomy**, vol. 9. 2019. https://doi.org/10.3390/agronomy9050262.

[7] Zhang, Y., Chen, H., Liang, Y., Lu, T., Liu, Z., Jin, X., Hou, L., Xu, J., Zhao, H., Shi, Y., and Ahammed, G. J. "Comparative transcriptomic and metabolomic analyses reveal the

protective effects of silicon against low phosphorus stress in tomato plants" **Plant Physiology and Biochemistry**, vol. 166. 2021. pp.78-87. https://doi.org/10.1016/j.plaphy.2021.05.043.

[8] Xu, X., Du, X., Wang, F., Sha, J., Chen, Q., Tian, G., Zhu, Z., Ge, S., and Jiang, Y. "Effects of Potassium Levels on Plant Growth, Accumulation and Distribution of Carbon, and Nitrate Metabolism in Apple Dwarf Rootstock Seedlings" **frontiers in plant science**, vol. 2020. https://doi.org/10.3389/fpls.2020.00904.

[9] Department of Agriculture. "Nutrient Management and Fertilizing Durian" [Online]. Available: https://www.doa.go.th/share/attachment.php?aid=2975. 2002.

[10] Chung, F. "Durian: Crop Production Cycle and Orchard Management Practices" [Online]. Available: http://durianinfo.blogspot.com/p/durian-crop-production-cycle.html. 2011.

[11] Amanah, H. Z., Joshi, R., Masithoh, R. E., Choung, M.-G., Kim, K.-H., Kim, G., and Cho, B.-K. "Nondestructive measurement of anthocyanin in intact soybean seed using Fourier Transform Near-Infrared (FT-NIR) and Fourier Transform Infrared (FT-IR) spectroscopy" **Infrared Physics & Technology**, vol. 111. 2020. pp.103477. https://doi.org/10.1016/j.infrared.2020.103477.

[12] Quintelas, C., Braga, A., Cordeiro, A., Ferreira, E. C., Belo, I., and Páscoa, R. N. M. J. "FT-NIR spectroscopy analysis for monitoring the microbial production of 2-phenylethanol using crude glycerol as carbon source" **LWT**, vol. 155. 2022. pp.112951. https://doi.org/10.1016/j.lwt.2021.112951.

[13] Laub-Ekgreen, M. H., Martinez-Lopez, B., Jessen, F., and Skov, T. "Non-destructive measurement of salt using NIR spectroscopy in the herring marinating process" **LWT**, vol. 97. 2018. pp.610-616. https://doi.org/10.1016/j.lwt.2018.07.024.

[14] Jin, X., Wang, L., Zheng, W., Zhang, X., Liu, L., Li, S., Rao, Y., and Xuan, J. "Predicting the nutrition deficiency of fresh pear leaves with a miniature near-infrared

spectrometer in the laboratory" **Measurement**, vol. 188. 2022. pp.110553. https://doi.org/10.1016/j.measurement.2021.110553.

[15] Sekerli, Y. E., Keskin, M., and Soysal, Y. "Testing of three sensor systems to predict water and nutrient contents of soccer field turfgrass clippings" **Urban Forestry & Urban Greening**, vol. 59. 2021. pp.126909. https://doi.org/10.1016/j.ufug.2020.126909.

[16] Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Ismail, M. F., Tan, N. P., and Karam, D. S. "Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest" **Computers and Electronics in Agriculture**, vol. 169. 2020. pp.105221. https://doi.org/10.1016/j.compag.2020.105221.

[17] Wang, S., Liu, S., Zhang, J., Che, X., Yuan, Y., Wang, Z., and Kong, D. "A new method of diesel fuel brands identification: SMOTE oversampling combined with XGBoost ensemble learning" **Fuel**, vol. 282. 2020. pp.118848. https://doi.org/10.1016/j.fuel.2020.118848.

[18] Au, J., Youngentob, K. N., Foley, W. J., Moore, B. D., and Fearn, T. "Sample selection, calibration and validation of models developed from a large dataset of near infrared spectra of tree leaves" **Journal of Near Infrared Spectroscopy**, vol. 28. 2020. pp.186-203. https://doi.org/10.1177/0967033520902536.

[19] Guo, P.-T., Li, M.-F., Luo, W., and Cha, Z.-Z. "Estimation of foliar nitrogen of rubber trees using hyperspectral reflectance with feature bands" **Infrared Physics & Technology**, vol. 102. 2019. pp.103021. https://doi.org/10.1016/j.infrared.2019.103021.

[20] Comino, F., Ayora-Cañada, M. J., Aranda, V., Díaz, A., and Domínguez-Vidal, A. "Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination" **Talanta**, vol. 188. 2018. pp.676-684. https://doi.org/10.1016/j.talanta.2018.06.058.

[21] Rébufa, C., Pany, I., and Bombarda, I. "NIR spectroscopy for the quality control of Moringa oleifera (Lam.) leaf powders: Prediction of minerals, protein and moisture contents" **Food Chemistry**, vol. 261. 2018. pp.311-321. https://doi.org/10.1016/j.foodchem.2018.04.066.

[22] Lequeue, G., Draye, X., and Baeten, V. "Determination by near infrared microscopy of the nitrogen and carbon content of tomato (Solanum lycopersicum L.) leaf powder" **Sci Rep**, vol. 6. 2016. pp.33183. https://doi.org/10.1038/srep33183.

[23] Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., and Yermiyahu, U. "Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance" **Biosystems Engineering**, vol. 114. 2013. pp.426-434. https://doi.org/10.1016/j.biosystemseng.2012.09.005.

[24] Zhai, Y., Cui, L., Zhou, X., Gao, Y., Fei, T., and Gao, W. "Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods" **International Journal of Remote Sensing**, vol. 34. 2012. pp.2502-2518. https://doi.org/10.1080/01431161.2012.746484.

[25] Ulissi, V., Antonucci, F., Benincasa, P., Farneselli, M., Tosti, G., Guiducci, M., Tei, F., Costa, C., Pallottino, F., Pari, L., and Menesatti, P. "Nitrogen concentration estimation in tomato leaves by VIS-NIR non-destructive spectroscopy" **Sensors (Basel)**, vol. 11. 2011. pp.6411-24. https://doi.org/10.3390/s110606411.

[26] Menesatti, P., Antonucci, F., Pallottino, F., Roccuzzo, G., Allegra, M., Stagno, F., and Intrigliolo, F. "Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [Citrus sinensis (L) Osbeck cv Tarocco]" **Biosystems Engineering**, vol. 105. 2010. pp.448-454. https://doi.org/10.1016/j.biosystemseng.2010.01.003.

[27] Richardson, A. D., Reeves Iii, J. B., and Gregoire, T. G. "Multivariate analyses of visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments" **New Phytologist**, vol. 161. 2003. pp.291-301. https://doi.org/10.1046/j.1469-8137.2003.00913.x.

[28] Wang, Y.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., and Zhang, Z.-Z. "NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves" **Infrared Physics & Technology**, vol. 108. 2020. pp.103365. https://doi.org/10.1016/j.infrared.2020.103365.

[29] Ciavarella, S. and Batten, G. D. "Measuring potassium in plant tissues using near Infrared spectroscopy" **Journal of Near Infrared Spectroscopy** vol. 6. 1998. https://doi.org/10.1255/jnirs.167.

[30] Chen, H., Lin, B., Cai, K., Chen, A., and Hong, S. "Quantitative analysis of organic acids in pomelo fruit using FT-NIR spectroscopy coupled with network kernel PLS regression" **Infrared Physics & Technology**, vol. 112. 2021. pp.103582. https://doi.org/10.1016/j.infrared.2020.103582.

[31] Jiménez-Carvelo, A. M., Martín-Torres, S., Ortega-Gavilán, F., and Camacho, J. "PLS-DA vs sparse PLS-DA in food traceability. A case study: Authentication of avocado samples" **Talanta**, vol. 224. 2021. pp.121904. https://doi.org/10.1016/j.talanta.2020.121904.

[32] Xia, Y. "Correlation and association analyses in microbiome study integrating multiomics in health and disease" **Prog Mol Biol Transl Sci**, vol. 171. 2020. pp.309-491. https://doi.org/10.1016/bs.pmbts.2020.04.003.

[33] Shekhar, S., Hoque, N., and Bhattacharyya, D. K. "PKNN-MIFS: A Parallel KNN Classifier over an Optimal Subset of Features" **Intelligent Systems with Applications**, vol. 14. 2022. pp.200073. https://doi.org/10.1016/j.iswa.2022.200073.

[34] Chen, Z., Zhao, F., Zhou, J., Huang, P., and Song, W. "A novel approach applied to fault diagnosis for micro-defects on piston throat" **Measurement**, vol. 173. 2021. pp.108508. https://doi.org/10.1016/j.measurement.2020.108508.

[35] Zahid, A., Dashtipour, K., Abbas, H. T., Mabrouk, I. B., Al-Hasan, M., Ren, A., Imran, M. A., Alomainy, A., and Abbasi, Q. H. "Machine learning enabled identification and real-time prediction of living plants' stress using terahertz waves" **Defence Technology**, vol. 2022. https://doi.org/10.1016/j.dt.2022.01.003.

[36] Choi, Y., Yoon, G., and Kim, J. "Unsupervised learning algorithm for signal validation in emergency situations at nuclear power plants" **Nuclear Engineering and Technology**, vol. 54. 2022. pp.1230-1244. https://doi.org/10.1016/j.net.2021.10.006.

[37] Aytaç Korkmaz, S. and Binol, H. "Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection" **Journal of Molecular Structure**, vol. 1156. 2018. pp.255-263. https://doi.org/10.1016/j.molstruc.2017.11.093.

[38] Bourquin, J., Schmidli, H., van Hoogevest, P., and Leuenberger, H. "Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form" **European Journal of Pharmaceutical Sciences**, vol. 7. 1998. pp.5-16. https://doi.org/10.1016/S0928-0987(97)10028-8.

[39] Shi, J., Wang, Y., Li, Z., Huang, X., Shen, T., and Zou, X. "Simultaneous and nondestructive diagnostics of nitrogen/magnesium/potassium-deficient cucumber leaf based on chlorophyll density distribution features" **Biosystems Engineering**, vol. 212. 2021. pp.458-467. https://doi.org/10.1016/j.biosystemseng.2021.11.001.

[40] James J, H. A., Wan I, W. I., Nawi, N. M., M. Shariff, A. R., and Mehdizadeh, S. A. "Application of Artificial Neural Network Classification to Determine Nutrient Content in Oil Palm Leaves" **Applied Engineering in Agriculture**, vol. 34. 2018. pp.497-504. https://doi.org/10.13031/aea.12403.

[41] Zeraatkar, S. and Afsari, F. "Interval–valued fuzzy and intuitionistic fuzzy–KNN for imbalanced data classification" **Expert Systems with Applications**, vol. 184. 2021. pp.115510. https://doi.org/10.1016/j.eswa.2021.115510.

[42] Liu, S., Wang, S., Hu, C., and Bi, W. "Determination of alcohols-diesel oil by near infrared spectroscopy based on gramian angular field image coding and deep learning" **Fuel**, vol. 309. 2022. pp.122121. https://doi.org/10.1016/j.fuel.2021.122121.

[43] Begum, N., Maiti, A., Chakravarty, D., and Das, B. S. "Diffuse reflectance spectroscopy based rapid coal rank estimation: A machine learning enabled framework" **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, vol. 263. 2021. pp.120150. https://doi.org/10.1016/j.saa.2021.120150.

[44] Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Tan, N. P., and Ismail, M. F. "Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and unmanned aerial vehicles" **Computers and Electronics in Agriculture**, vol. 193. 2022. pp.106646. https://doi.org/10.1016/j.compag.2021.106646.

[45] Sun, Y., Liu, N., Kang, X., Zhao, Y., Cao, R., Ning, J., Ding, H., Sheng, X., and Zhou, D. "Rapid identification of geographical origin of sea cucumbers Apostichopus japonicus using FT-NIR coupled with light gradient boosting machine" **Food Control**, vol. 124. 2021. pp.107883. https://doi.org/10.1016/j.foodcont.2021.107883.

[46] Xie, X.-L. and Li, A.-B. "Identification of soil profile classes using depth-weighted visible–near-infrared spectral reflectance" **Geoderma**, vol. 325. 2018. pp.90-101. https://doi.org/10.1016/j.geoderma.2018.03.029.

[47] Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Overall precision test for determination the nutrient in durian leaf in durian orchard using near-infrared spectroscopy" **Engineering and Applied Science Research**, vol. 49. 2022. https://doi.org/10.14456/easr.2022.15.

[48] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, P. W. "SMOTE: Synthetic Minority Over-sampling Technique" **ournal of Artificial Intelligence Research**, vol. 16. 2002. pp.321-357.

[49] Brownlee, J. "SMOTE for Imbalanced Classification with Python" [Online]. Available: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification. 2020.

[50] Chemchem, A., Alin, F., and Krajecki, M. "Combining SMOTE sampling and Machine Learning forForecasting Wheat Yields in France," presented at the 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy, June 3-5, 2020.

[51] Xu, L., Zhou, Y.-P., Tang, L.-J., Wu, H.-L., Jiang, J.-H., Shen, G.-L., and Yu, R.-Q. "Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration" **Analytica Chimica Acta**, vol. 616. 2008. pp.138-143. https://doi.org/10.1016/j.aca.2008.04.031.

[52] Shafizadeh-Moghadam, H. "Fully component selection: An efficient combination of feature selection and principal component analysis to increase model performance" **Expert Systems with Applications**, vol. 186. 2021. pp.115678. https://doi.org/10.1016/j.eswa.2021.115678.

[53] Lindgren, I. "Dealing with Highly Dimensional Data using Principal Component Analysis (PCA)" [Online]. Available: https://towardsdatascience.com/dealing-with-highly-dimensional-data-using-principal-component-analysis-pca-fea1ca817fe6. 2020.

[54] Gilmer-Osborne, B., Fearn, T., Hindle, P. H., and Hindle, P. T. **Practical Nir Spectroscopy With Applications in Food and Beverage Analysis.** UK: Longman group. 1993.

[55] Land Development Department. "Soil management information" [Online]. Available: https://www.ldd.go.th/Web_Soil/acid.htm. 2020.

[56] Milali, M. P., Sikulu-Lord, M. T., Kiware, S. S., Dowell, F. E., Corliss, G. F., and Povinelli, R. J. "Age grading An. gambiae and An. arabiensis using near infrared spectra and artificial neural networks" **PLoS One**, vol. 14. 2019. pp.e0209451. https://doi.org/10.1371/journal.pone.0209451.

[57] Min, M., Lee, W. S., Kim, Y. H., and Bucklin, R. A. "Nondestructive Detection of Nitrogen in Chinese Cabbage Leaves UsingVIS–NIR Spectroscopy" **HortScience**, vol. 41. 2006. pp.162-166. https://doi.org/10.21273/HORTSCI.41.1.162.

[58] Grimm, B. **Advances in botanical research.** Elsevier Ltd. 2020.

[59] Heldt, H.-W. and Piechulla, B.: Academic Press. 2021.

[60] Lafi, S. Q. and Kaneene, J. B. "An explanation of the use of principal-components analysis to detect and correct for multicollinearity" **Preventive Veterinary Medicine**, vol. 13. 1992. pp.261-275. https://doi.org/10.1016/0167-5877(92)90041-D.

[61] Manolopoulou, E. and Varzakas, T. "Effect Of Temperature In Color Changes Of Green Vegetables" **Current Research in Nutrition and Food Science Journal**, vol. 4. 2016. pp.10-17. 10.12944/CRNFSJ.4.Special-Issue-October.02.

[62] Fiorentini, R. and Galoppini, C. "The proteins from leaves" **Plant Foods for Human Nutrition**, vol. 32. 1983. pp.335-350.

[63] Wong, T.-T. and Tsai, H.-C. "Multinomial naïve Bayesian classifier with generalized Dirichlet priors for high-dimensional imbalanced data" **Knowledge-Based Systems**, vol. 228. 2021. pp.107288. https://doi.org/10.1016/j.knosys.2021.107288.

[64] Xiao, Y., Wu, J., and Lin, Z. "Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data" **Computers in Biology and Medicine**, vol. 135. 2021. pp.104540. https://doi.org/10.1016/j.compbiomed.2021.104540.

[65] Phy, V. "Accuracy is NOT enough for Classification Tasks" vol. 2019. [Online]. Available: https://towardsdatascience.com/accuracy-is-not-enough-for-classification-task-47fca7d6a8ec

[66] Brownlee, J. "A Gentle Introduction to Imbalanced Classification" [Online]. Available: https://machinelearningmastery.com/what-is-imbalanced-classification/. 2020.

[67] Bae, S.-Y., Lee, J., Jeong, J., Lim, C., and Choi, J. "Effective data-balancing methods for class-imbalanced genotoxicity datasets using machine learning algorithms and molecular fingerprints" **Computational Toxicology**, vol. 20. 2021. pp.100178. https://doi.org/10.1016/j.comtox.2021.100178.

Chapter 5

Primary nutrients assessment of durian (CV Monthong) leaf sample matrixes using near infrared spectroscopy with wavelength selection

Farmers would be able to regulate fertilization and produce quality durian if they knew the nutrient concentration in durian leaves. A long period of time for traditional nutritional content determination is needed. Therefore, near-infrared spectroscopy is a good method for nondestructive and quick nutrient content evaluation. The leaf sample matrices (fresh leaf, dried ground leaves, and dried ground leaf pellet) were scanned by Fourier transform near-infrared (FT-NIR) with a wavelength of 12,500-3,600 cm^{-1}. Regression models were developed using partial least squares (PLS) with full wavelength, short wavelength (7,500-6,000 cm^{-1} + 6,600-4,000 cm^{-1} for fresh leaf spectra and 6,000-4,000 cm^{-1} for dried ground leaves and dried ground leaves pellet spectra), and selected wavelength by successive projections algorithm (SPA). In this study, the model for N and K concentration was acceptable and the prediction was considered good but for P content not had succeeded. As a result, the PLS-SPA model using fresh leaf samples for evaluating N content in durian leaves exhibited performance of r^2 = 0.852, SEP = 0.14%, RPD = 2.63 and bias = -0.020%. The PLS-SPA model using dried ground leaves samples for evaluating K content in durian leaves exhibited performance of r^2 = 0.820, SEP = 0.13%, RPD = 2.36 and bias = 0.006%.

* This chapter is under reviewed by Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy journal.

5.1 Introduction

Nutrients are required for horticulture to grow properly and provide high-quality products. Nitrogen (N), phosphorus (P), and potassium (K), which are primary macronutrients, are favorable to crop quality or growth [1]. Plant nutrient deficiency causes a slowing growth rate and low production yields [2]. In addition, overfertilization may result in undesirable crop productivity and quality development [3]. Fertilizer management is crucial for horticulture so that plants receive adequate fertilizer without overfertilization and receive the most benefit. Meanwhile, farmers can save on fertilizer costs. In Thailand, the exports of fresh durian, frozen durian whole fruit, freeze-dried durian, and processed durian products are increasing every year [4]. As a result, farmers are especially confronted with producing large quantity and high-quality crops because durian is a high-value fruit.

Information on fertilizer management based on durian growth stages can support plant growth and yield because each stage necessitates a different amount of nutrients. There are 4 stages of durian cultivation: the unproductive stage (the first 1-4 years) and every year or season of the flowering stage, the fruiting stage, and the harvesting stage. In each stage, durian trees need nutrient according to Phanomsophon, 2022 [5]. In durian leaves, the suitable nutritional content range is N = 2.0-2.3%, P = 0.15-0.25%, and K = 1.7-2.5% [6].

Normally, the nutritional content of the leaf can be determined by chemicals, which requires a long period of time for measurement. Near infrared (NIR) spectroscopy is a rapid alternative technique that helps farmers evaluate the concentration of nutrients immediately [7] with a one-minute NIR measurement. Fourier transform near-infrared (FT-NIR) is a fixed NIR measurement that covers NIR bands (10,000-3,600 cm^{-1}) [8]. When NIR light passes through a sample, the C-H, O-H, and N-H bonds inside the material absorb some energy, causing molecular vibrations that are related to the sample's molecular

structure [9]. Although -P and -K bonds are unable to absorb energy in the NIR region, several studies have successfully used NIR spectroscopy to estimate P and K concentrations in leaves. The P and K concentrations are related to proteins, chlorophyll, and carbohydrates that can absorb NIR waves [10]. P is found in nucleic acids and nucleoproteins, while K is important for photosynthesis and carbohydrate metabolism [10]. Previous research has used NIR spectroscopy and chemometrics for modeling leaf samples to quantify N [11-15], P [10, 15-18], and K concentrations [10, 12, 19-22]. To evaluate nutritional content in plants, the studies employed fresh leaf samples or dried ground leaf samples. Model performance is influenced by sample type because of the spectral characteristics induced by surface conditions, particle size, and moisture effects [23]. In addition to the type of sample, the wavelength selection also affects the performance of the model. It is possible to increase model performance for nutrient evaluation by selecting wavelengths suitable to the macronutrient concentration or by deleting undesirable wavelengths [24, 25].

This research focused on the influence of the leaf sample matrices (fresh leaf, dried ground leaves, and dried ground leaf pellet) on the model performance (coefficient of determination of prediction set (r^2) and ratio of prediction to deviation (RPD)) of nutrient evaluation in durian (CV Monthong) leaf and on the adjustment of the relevant variable inputs for estimating a nutrient quantity using partial least squares (PLS) regression with spectrum pretreatment and wavelength selection (full wavelength, short wavelength, and selected wavelength by successive projections algorithm (SPA). The goals of this study were as follows: 1) to rapidly measure macronutrient concentrations using an FT-NIR spectrometer with PLS; 2) to observe the wavelength selection suitable for macronutrient content prediction; and 3) to determine the suitable sample matrix type for macronutrient content prediction. There has been no research on using NIR to determine the nutrient content of durian (CV Monthong) leaves; however, the classification of nutrient content levels has been reported by Phanomsophon et al. [5]. If this approach is successful, it can

assist farmers in controlling durian tree fertilization, lowering expenses and increasing revenues.

5.2 Materials and Methods

5.2.1 Durian leaf sample matrices

This experimental sample was collected in a durian orchard in eastern Thailand, including *Durio zibethinus* Murray CV Monthong orchards. Two orchards from Chanthaburi (12° 51' 29" N and 102° 9' 16" E), two orchards from Rayong (12° 49' 60" N and 101° 25' 60" E), and two orchards from Trad (12° 18' 09" N and 102° 30' 45" E) were chosen for this study. Each orchard's durian tree included in the experiment was older than eight years according to Phuwarodom [6]. From December 2018 to November 2020, durian leaf samples were collected in the flowering, fruiting and after harvesting stages of annual durian production and utilized for modeling.

One hundred and eleven samples were obtained in total (one sample consisted of 20 fresh durian leaves). The fresh leaf samples (1) and dried ground leaf sample (2) were prepared according to Phanomsophon, 2022 [5]. Each dried ground leaf pellet sample (3) was prepared by using 0.15±0.01 g dried ground leaves placed in a stainless block (diameter 1.40 cm) and pressurized at 60 MPa (Figure 5.1).

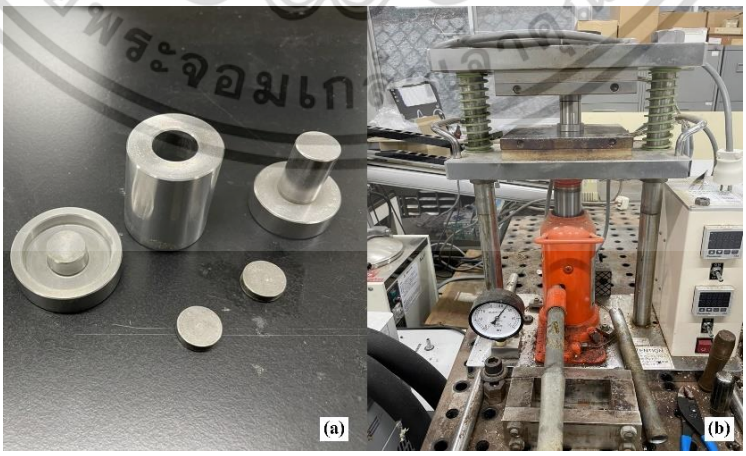


Figure 5.1 The stainless block (a) and pressure machine (b) for pelleting.

5.2.2 NIR measurement

The NIR scanning according to Phanomsophon, 2022 [5]. An FT-NIR spectrometer (Bruker Ltd., Germany) was used to scan the samples in reflectance mode with a wavelength range of 12,500-3,500 cm^{-1}, an average of 32 scans and a resolution of 32 cm^{-1}. Gold was used as the reference material. The scanning was performed under room conditions at 25±2°C. Figure 5.2 shows the NIR scanning. Before scanning, the fresh leaf rested for 30 minutes in an ice bucket. The fresh leaf was scanned under an aluminum plate. One fresh leaf spectrum was representative of a tree and was obtained by averaging the spectra of 20 fresh leaves and 2 scan positions per leaf. A quartz cup (diameter 4.5 cm and height of 5.3 cm) was used to scan the dried ground leaf sample. The dried ground leaf pellet sample was covered with Spectralon reference material and covered with an aluminum can before scanning because the sample was smaller than the scan window. This must be done to avoid the problem of interference between light from outside and of light leaked from a source. The process time of flesh leaves (FL) was 20 minutes (in total 20 leaves), while it was 1 minute for dried ground leaves (DL) and dried ground leaf pellets (PL).

5.2.3 Nutrient measurement

Nutrient analysis according to Phanomsophon, 2022 [5]. Only dried ground leaf samples were used to measure the primary macronutrients. The nitrogen (N) content was measured with 0.10 g by TruMac CNS (Leco, USA). Before measurement, the samples were dried at 70°C for 48 hr. Each sample was prepared for phosphorus (P) and potassium (K) concentration measurements by heating at 500°C for 6 hrs with 0.25 g, digesting with 10 ml of aqua regia (Merck, Germany), and measuring with ICP-OES (Perkin Elmer, USA).

.

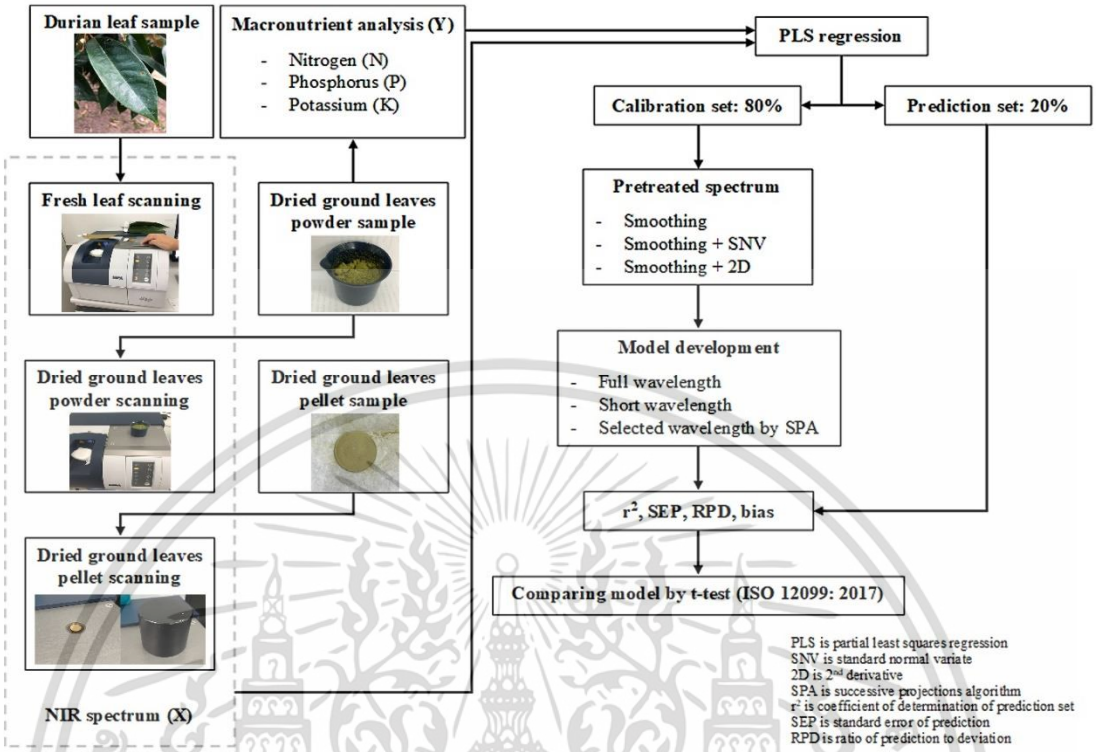


Figure 5.2 Predictive macronutrient concentration modeling flow charts.

5.2.4 Model development

In this work, the models were developed by MATLAB R2022b software (Mathwork, USA). The spectrum used for model development consisted of both nontreatment and pretreatment, which included 1) raw spectra, 2) smoothing spectra (SMT), 3) smoothing + standard normal variate (SMT+SNV) and 4) smoothing +2nd derivative (SMT+2D). The wavelengths used for modeling were the full wavelength (12,500-3,600 cm^{-1}), short wavelength range (selected by high absorption), and wavelength selected by the successive projection algorithm (SPA). In the case of the short wavelength range were chosen from the peak that appears clearly on the spectrum, the selected ranges for the raw spectra, SMT, SMT+SNV and SMT+2D of fresh leaves (FL) were 7,200-4,400 cm^{-1}, 7,200-4,400 cm^{-1}, 8,000-4,000 cm^{-1} and 7,500-6,600 cm^{-1} + 6,000-4,000 cm^{-1}, respectively. For dried ground leaves (DL) and dried ground leaf pellets (PL), wavelength ranges of 5,400-

4,000 cm^{-1}, 5,400-4,000 cm^{-1}, 8,000-4,000 cm^{-1} and 6,000-4,000 cm^{-1} were selected for raw, SMT, SMT+SNV and SMT+2D spectra, respectively.

Wavelength selection is a crucial factor for modeling because it enhances the prediction performance, decreases complexity, and produces a personalized spectrometer with a high signal-to-noise ratio [26]. SPA is a popular method for wavelength selection that reduces the number of wavelengths by minimizing collinearity between variables [27]. SPA is a forward feature selection technique that chooses one wavelength and then chooses a new wavelength with the maximum projection value on the orthogonal subspace of the previously chosen wavelength, and the procedure is repeated for all wavelengths [28].

Partial least squares (PLS) regression was used to create the model using leave-one-out cross-validation. The PLS calibration principle is to maximize the covariance between the matrix X (spectrum) and the response vector Y (reference value), which causes the latent variables (LVs) to be ranked in terms of their importance in explaining X and Y [29]. The number of LVs has an impact on model performance. More LVs increase the danger of incorporating more noise than information [30].

The model can perform well if the proper parameters are chosen. The number of wavelengths (nWL) and number of LVs (nLVs) were the major parameters for PLS modeling on which this work focused. Figure 5.3 shows the variation in SECV/SEP and R^2/r^2 with nWL (a) and with nLVs (b). The elbow point approach was used to find the proper nWL and wLVs. This decrease/increase begins to flatten evidently at some nWL and nLVs, which is referred to as the 'elbow' point. The elbow method focuses on monotonically decreasing SECV/SEP and monotonically increasing R^2/r^2 as nWL and nLVs increase [31].

.

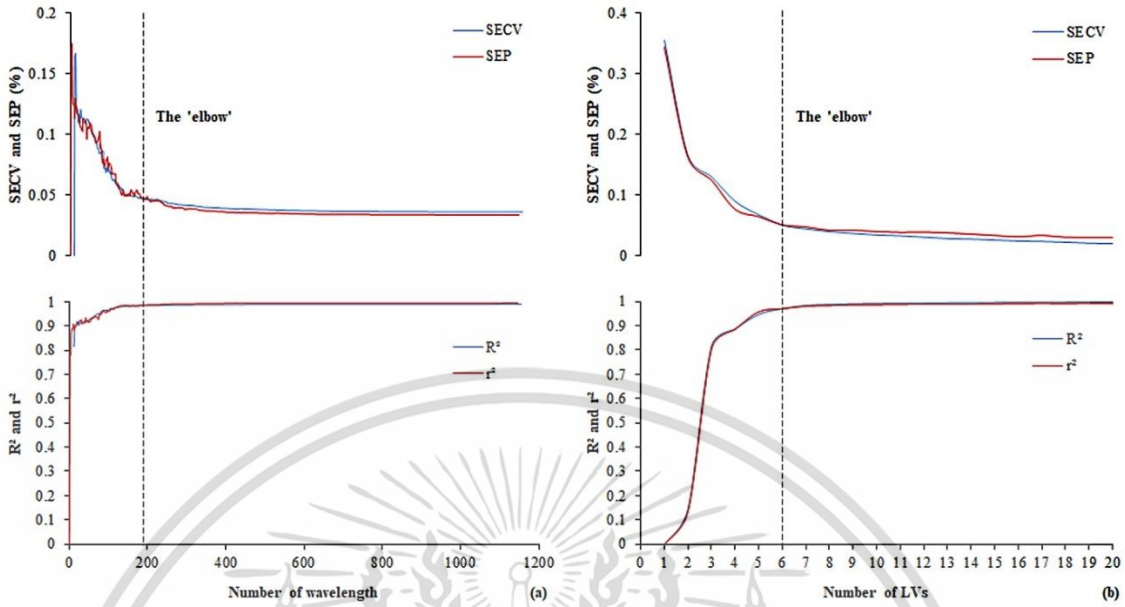


Figure 5.3 Trend of SECV/SEP and R^2/r^2 versus increasing number of wavelengths (a) and LVs (b).

5.2.5 Model performance

In this work, the performance was calculated by MATLAB R2022b software (Mathwork, USA). To estimate the performance of the model, the models were validated by cross validation, and the following indices were calculated: for calibration set, the coefficient of determination of the cross-validation set (R^2), the standard deviation of error of cross-validation (SECV) were considered and for prediction set, the coefficient of determination of the prediction set (r^2), the standard deviation of error of prediction (SEP), bias, and the ratio of prediction to deviation (RPD) were considered, which were calculated by the following formulas:

$$R^2, r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5.1)$$

$$SECV, SEP = \sqrt{\frac{\sum \left((y_i - \hat{y}_i) - \frac{\sum (y_i - \hat{y}_i)}{n} \right)^2}{n-1}} \quad (5.2)$$

$$\text{Bias} = \frac{\sum y_i - \hat{y}_i}{n} \quad (5.3)$$

$$\text{RPD} = \frac{SD_y}{SEP} \quad (5.4)$$

where y_i is the reference value

$\hat{y}_i$ is the predicted value

$\bar{y}$ is the average of the reference value

n is the number of samples

SD_y is the standard deviation of the reference value of the prediction set. Following the selection of the models, the r^2 and RPD values were used to assess the model's performance [32].

5.2.6 Significance evaluation of the model by t test (ISO 12099)

According to ISO 12099:2017, the determination of the statistics checking of bias, SEP and slope by t test can be conducted [33]. The bias can be checked with the bias confidence limits (T_b). If the bias value is less than T_b, the bias is not significantly different from zero, indicating a 0% prediction error. T_b can be calculated by the following formula:

$$T_b = \pm \frac{t_{(1-\alpha/2)} SEP}{\sqrt{n}} \quad (5.5)$$

where $t_{(1-\alpha/2)}$ is the appropriate Student's t value for a two-tailed test with degrees of freedom associated with SEP and the selected probability of a type I error, α is the probability of making a type I error

n is the number of independent samples

The SEP can be checked with the unexplained error confidence limit (T_{UE}). If the SEP value is less than the T_{UE} value, the SEP can be accepted, indicating the high precision of the prediction model. T_{UE} can be calculated using the following formula:

$$T_{UE} = SEC\sqrt{F_{(\alpha,v,M)}} \quad (5.6)$$

where F is the F test value

v = n – 1 is the numerator degree of freedom associated with the SEP of the test set, in which n is the number of samples in the validation process

M = n_c – p – 1 is the denominator degree of freedom associated with the SEC, in which n_c is the number of calibration samples and p is the number of LVs.

The slope of the scatter plot between the reference value and predicted value can be checked with the observed t value (t_{obs}). If t_{obs} is less than $t_{(1-\alpha/2)}$, the slope is not different from 1 (b = 1), indicating 100% prediction accuracy of the model. t_{obs} can be calculated using the following formula:

$$t_{obs} = |b-1| \sqrt{\frac{S_{\hat{y}}^2 (n-1)}{S_{res}^2}} \quad (5.7)$$

$$S_{res} = \sqrt{\frac{\sum [y_i - (a+b\hat{y}_i)]^2}{n-2}} \quad (5.8)$$

where a is the intercept of the simple regression

b is slope of the simple regression

$S_{\hat{y}}^2$ is the variance of the n predicted values

S_{res} is the residual standard deviation.

This study applied statistical checking of the slope with the predicted value from the model and another model to check the hypothesis that b =1. If t_{obs} is less than $t_{(1-\alpha/2)}$, the predicted value from the model does not differ from another model.

5.3 Results

5.3.1 Data on the macronutrient content of durian leaf samples

All macronutrient content data (N, P and K) were sorted ascending and split to 80%:20% for the calibration set to the prediction set. The calibration set range was covered by the prediction set range. Based on 111 samples, it was possible to separate

them into 89 calibration samples and 22 prediction samples, which were the same for fresh leaf (FL), dried ground leaf (DL), and dried ground leaf pellet samples (PL). The statistical characteristics of all nutritional data were normally distributed. The concentrations of macronutrients in durian leaf samples were N = 1.53-3.87%, P = 0.05-0.39%, and K = 1.04-2.44% (Table 5.1).

Table 5.1 Statistical characteristics of macronutrient data.

Parameters	Datasets	N	Max (%)	Mean (%)	Min (%)	SD (%)
N	Calibration	89	3.87	2.50	1.53	0.39
	Prediction	22	3.48	2.51	1.81	0.37
P	Calibration	89	0.39	0.22	0.05	0.05
	Prediction	22	0.32	0.22	0.11	0.05
K	Calibration	89	2.44	1.61	1.04	0.30
	Prediction	22	2.37	1.62	1.10	0.30

*N: number of samples, SD: the standard deviation.

5.3.2 NIR spectrum of durian leaf sample matrices

The raw spectrum of FL samples (Figure 5.4a) showed high absorption at 6881 cm^{-1} and 5176 cm^{-1} (i.e., vibration bands of sucrose, starch, and H_2O [9], respectively). The raw spectra of the DL and PL samples (Figure 5.4a) had the same NIR absorption behavior; in both spectra, only the baseline shift appeared due to the different densities. The spectra were pretreated with smoothing (SMT), smoothing + standard normal variate (SMT+SNV) and smoothing +2nd derivative (SMT+2D), and it was found that the correlation between macronutrient content and absorption in some wavelengths of SMT+2D spectra was higher than that of other spectra. In Figure 5.4b, the SMT+2D spectra of FL samples showed high absorption at 7212, 7097, 5353, 5276, 4489 and 4188 cm^{-1} (under zero-line peaks), which correspond to the chemical bands of CH_2, ROH, starch, starch, amino acid and cellulose, respectively [9]). For the DL and PL samples (Figure 5.4b), high peaks were observed at

5276, 4489 cm^{-1} and 4188 cm^{-1} (under zero-line peaks). These absorbance values are the vibration bands of starch, amino acids and cellulose, respectively [9]. The indication of water in FL was discarded in the DL and PL samples, and N, P and K are not atoms in the H_2O molecule.

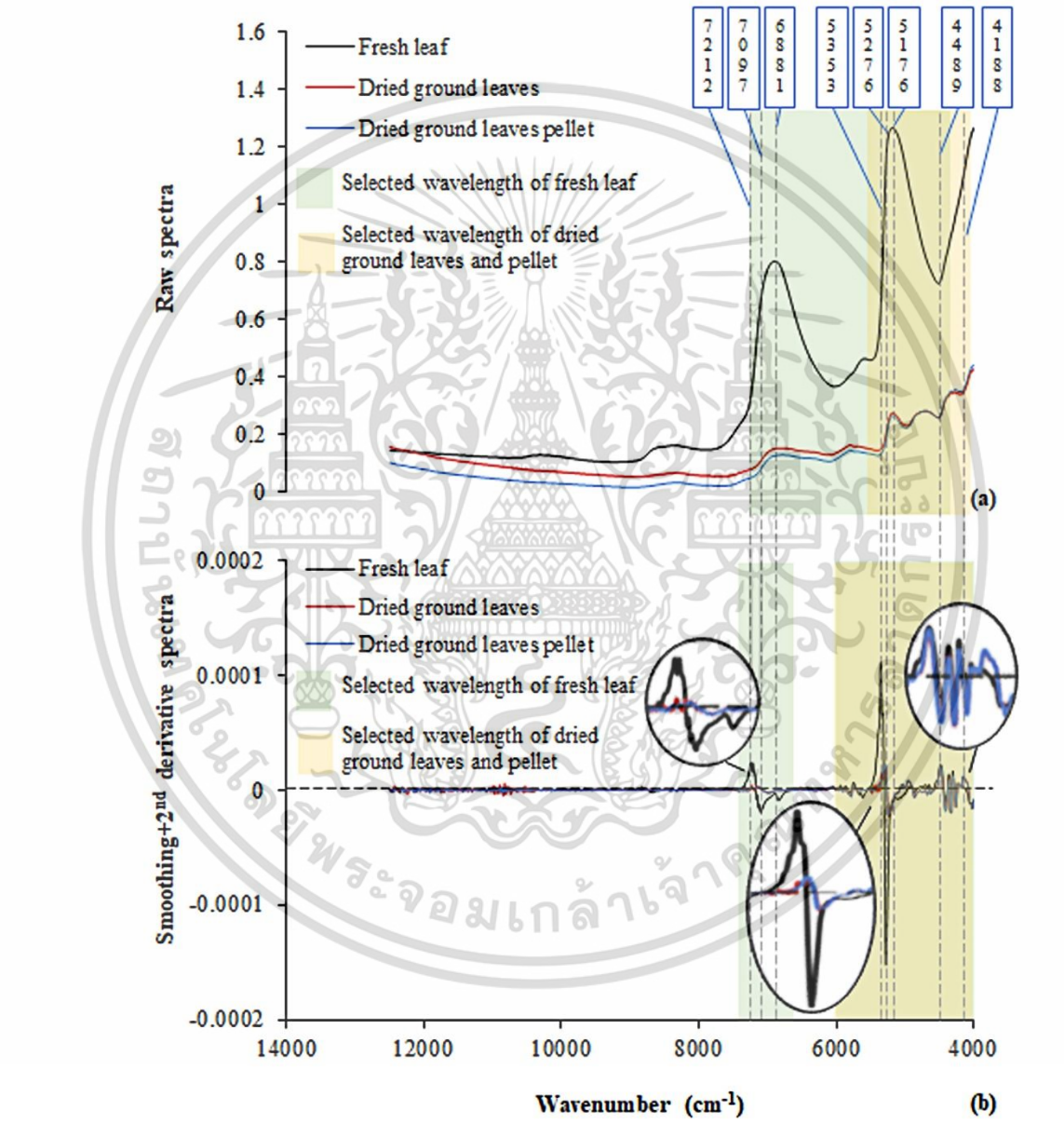


Figure 5.4 Raw spectra (a) and smoothing+2nd derivative spectra (b) of samples.

5.3.3 Comparison of predicted macronutrient concentration models for durian leaf sample matrices

The type of sample, number of wavelengths, spectral pretreatment, and number of PLS factors are all aspects that influenced the performance of the PLS regression model. Spectra pretreatment including smoothing (SMT) spectra, smoothing + standard normal variate (SMT+SNV) and smoothing +2nd derivative (SMT+2D) could boost performance of models. Table 5.2 revealed that the model with a wavelength selection had slightly different r^2 values to using the full wavelength, and in some cases, the model performed even better. Unwanted noise was decreased by removing wavelengths that were unrelated to nutrient content [34]. A comparison of model performance for sample matrices showed that the models for N content prediction from FL, DL, or PL had similar r^2 values. In models for K content prediction, the use of DL and PL spectra yielded slightly higher r^2 values compared to models using FL spectra. However, r^2 values were poor when predicting P due to its very small amount in the leaf (<0.40%).

The most accurate model for estimating N content in durian leaves was created using full-SMT spectra of DL samples with LVs = 14, which achieved an r^2 value of 0.912, SEP of 0.11%, RPD of 3.45, and bias of -0.025%. This model was considered excellent (r^2 > 0.90), and the prediction quality was rated as good to exceptional (RPD > 3.0) [35]. A comparison of the predicted values from the best model and other models of the same sample type was conducted using a t-test to check the hypothesis of the slope of the scatter plot. For FL sample, the PLS-full wavelength, PLS-short wavelength, and PLS-SPA models showed no significant difference compared to the best model at a 95% confidence interval. These results indicated similar accuracy of prediction. Similarly, for DL sample, the models showed no significant difference, except for the PLS-SPA model with raw spectrum, which showed a significant difference compared to the others. Regarding the PL models, the PLS-short wavelength and PLS-SPA models showed no significant difference, but the PLS-full wavelength model exhibited a significant difference.

The results were insufficient for predicting P content in durian leaf. None of the models had a r^2 value more than 0.66 and RPD greater than 2.00 [35] in matrix durian leaf samples.

The DL sample with LVs = 14 exhibited the highest accuracy in estimating the K content in durian leaves, as indicated by an r^2 value of 0.820, SEP of 0.13%, RPD of 2.36, and bias of 0.006%. The models displayed r^2 values ranging from 0.80 to 0.90, indicating good prediction accuracy. Additionally, the RPD values ranged from 2.00 to 2.50, demonstrating that the models were capable of providing approximate quantitative predictions [35].

To test the slope hypothesis, a t-test was conducted to compare the predicted values from the best model and other models of the same sample type for K content estimating. The results showed that the models of the FL sample were not significantly different from the best model at the 95% confidence interval. However, the PLS-short wavelength with SMT+SNV spectra and PLS-SPA with SMT spectra exhibited no significant difference in the DL sample, while other models (PLS-full with raw spectra, PLS-full with SMT spectra, PLS-short wavelength with raw spectra and PLS-SPA with raw spectra) showed considerable differences compared to the best model (PLS-short wavelength with SMT+SNV spectra). In the PL sample, PLS-full wavelength with raw spectra and SMT spectra did not show significant differences, as well as PLS-SPA with SMT+SNV spectra. However, the other models were significantly different from the best model.

The slope hypothesis between the best model and other models in the same sample type was tested by t test, and it was found that they were not different in the prediction of some model of N and K which were predicted in durian leaves. The P models were not tested because of poorly.

Table 5.2 The model prediction results for macronutrient concentrations in durian leaf sample matrices.

Nutrient	Sample	Wave-lengths	Pretreatment	Calibration set				Prediction set				
				nLVs	nWL	R^2	SECV (%)	r^2	SEP (%)	RPD	Bias (%)	Slope t test
N	Fresh leaf	Full	raw	12	1154	0.856	0.15	0.819	0.16	2.35	-0.003	NS
			SMT	12	1154	0.845	0.15	0.815	0.16	2.33	0.000	NS
		Short	raw	8	365	0.806	0.17	0.745	0.19	2.00	-0.023	NS
			SMT	15	365	0.885	0.13	0.851	0.14	2.60	0.002	NS
		SPA	raw	10	200	0.780	0.18	0.605	0.23	1.60	-0.023	NS
			SMT+2D	**15**	**286**	**0.900**	**0.12**	**0.852**	**0.14**	**2.63**	**-0.020**	*****
	Dried ground leaves	Full	raw	14	1154	0.966	0.07	0.874	0.13	2.86	-0.021	NS
			SMT	14	1154	0.948	0.09	0.912	0.11	3.45	-0.025	*
		Short	raw	12	182	0.903	0.12	0.852	0.14	2.72	0.042	NS
			SMT+SNV	14	182	0.902	0.12	0.869	0.13	2.80	0.021	NS
		SPA	raw	11	155	0.786	0.18	0.638	0.22	1.66	-0.002	S
			SMT	**11**	**235**	**0.891**	**0.13**	**0.879**	**0.13**	**2.89**	**0.011**	**NS**
	Dried ground leaves pellet	Full	raw	7	1154	0.809	0.17	0.749	0.18	2.00	0.011	S
			SMT	8	1154	0.838	0.15	0.786	0.17	2.16	-0.001	S
		Short	raw	8	182	0.877	0.14	0.819	0.14	2.57	0.062	NS
			SMT+SNV	12	182	0.909	0.12	0.851	0.12	2.98	0.069	NS
		SPA	raw	10	121	0.771	0.18	0.718	0.18	2.01	0.067	NS
			SMT	**11**	**235**	**0.887**	**0.13**	**0.880**	**0.12**	**3.10**	**0.045**	*****

Nutrient	Sample	Wave-lengths	Pretreatment	Calibration set				Prediction set				
				nLVs	nWL	R^2	SECV (%)	r^2	SEP (%)	RPD	Bias (%)	Slope t test
P	Fresh leaf	Full	raw	12	1154	0.620	0.03	0.118	0.04	1.06	0.002	S
			SMT	3	1154	0.154	0.05	0.232	0.04	1.14	0.02	S
		Short	raw	10	365	0.497	0.04	0.019	0.05	1.03	0.008	S
			SMT+2D	4	378	0.283	0.04	0.235	0.04	1.14	0.001	NS
		SPA	raw	12	57	0.371	0.04	0.140	0.04	1.11	-0.011	S
			SMT+2D	**5**	**372**	**0.327**	**0.04**	**0.323**	**0.04**	**1.22**	**0.000**	*
	Dried ground leaves	Full	raw	4	1154	0.228	0.05	0.029	0.05	1.00	-0.009	S
			SMT+2D	3	1154	0.411	0.04	0.276	0.04	1.18	0.000	NS
		Short	raw	3	182	0.191	0.05	0.055	0.05	1.03	0.001	S
			SMT+2D	6	260	0.557	0.04	0.100	0.04	1.08	-0.009	NS
		SPA	raw	13	15	0.395	0.04	0.137	0.04	1.08	0.001	NS
			SMT+2D	**3**	**361**	**0.447**	**0.04**	**0.341**	**0.04**	**1.23**	**0.003**	*
	Dried ground leaves pellet	Full	raw	3	1154	0.143	0.05	0.121	0.05	0.95	-0.003	NS
			SMT+SNV	9	1154	0.523	0.04	0.258	0.04	1.16	-0.001	NS
		Short	raw	5	182	0.254	0.05	0.086	0.04	1.06	0.008	S
			SMT+2D	**5**	**260**	**0.485**	**0.04**	**0.362**	**0.04**	**1.26**	**-0.003**	*
		SPA	raw	8	123	0.311	0.04	0.120	0.04	1.08	0.006	S
			SMT+2D	6	150	0.393	0.04	0.220	0.04	1.13	0.003	S

Nutrient	Sample	Wave-lengths	Pretreatment	Calibration set				Prediction set				
				nLVs	nWL	R^2	SECV (%)	r^2	SEP (%)	RPD	Bias (%)	Slope t test
K	Fresh leaf	Full	raw	13	1154	0.811	0.13	0.383	0.24	1.27	0.008	NS
			SMT+2D	8	1154	0.849	0.12	0.678	0.17	1.81	0.040	*
		Short	raw	5	365	0.361	0.24	0.388	0.23	1.33	0.065	NS
			SMT+SNV	10	365	0.602	0.19	0.524	0.20	1.51	0.059	NS
		SPA	raw	15	202	0.762	0.14	0.678	0.17	1.76	-0.009	NS
			SMT+2D	**14**	**45**	**0.670**	**0.17**	**0.657**	**0.17**	**1.76**	**0.043**	**NS**
	Dried ground leaves	Full	raw	10	1154	0.803	0.13	0.715	0.15	1.98	0.052	S
			SMT	10	1154	0.765	0.14	0.758	0.14	2.13	0.043	S
		Short	raw	8	182	0.607	0.19	0.597	0.19	1.63	0.050	S
			SMT+SNV	18	182	0.825	0.12	0.834	0.12	2.54	0.030	*
		SPA	raw	10	171	0.647	0.18	0.642	0.18	1.72	0.044	S
			SMT	**14**	**234**	**0.811**	**0.13**	**0.820**	**0.13**	**2.36**	**0.006**	**NS**
	Dried ground leaves pellet	Full	raw	13	1154	0.849	0.12	0.810	0.13	2.34	-0.025	NS
			SMT	13	1154	0.817	0.13	0.836	0.12	2.51	-0.021	*
		Short	raw	18	182	0.858	0.11	0.783	0.14	2.19	-0.028	S
			SMT+SNV	19	182	0.814	0.13	0.747	0.15	2.01	-0.022	S
		SPA	raw	10	42	0.675	0.17	0.400	0.23	1.31	0.042	S
			SMT+SNV	**11**	**971**	**0.800**	**0.13**	**0.808**	**0.13**	**2.29**	**-0.013**	**NS**

* nLVs, number of latent variables (PLS factors); nWL, number of wavelengths; R^2, coefficient of determination of calibration set; SECV, standard error of cross validation; r^2, coefficient of determination of prediction set; SEP, standard error of prediction; RPD, ratio of prediction to deviation; SPA, successive projections algorithm; SMT, smoothing; SMT+SNV, smoothing+standard normal variate; SMT+2D, smoothing+2nd derivative; *, the highest r^2 model; NS, no significant difference between NIR-predicted macronutrient concentration of the highest r^2 model and itself model at 95% confidence interval; S, significant difference between NIR-predicted nutrient concentration of the highest r^2 model and itself model at 95% confidence interval.

The models with high r^2, least number of wavelengths, least number of LVs and no significant difference with the best model were suggested as a representation for macronutrient prediction in durian leaves because this method was fastest, and it used only important wavelengths indicating for test in prediction. Thus, the performance of the selected models had to be further evaluated with a bias test, SEP test and slope test for whether SEP and bias can be considered to be zero or not, and slope can be one or not, according to ISO 12099.

Table 5.3 The performance measurement statistics by t test of selected models followed in ISO12099.

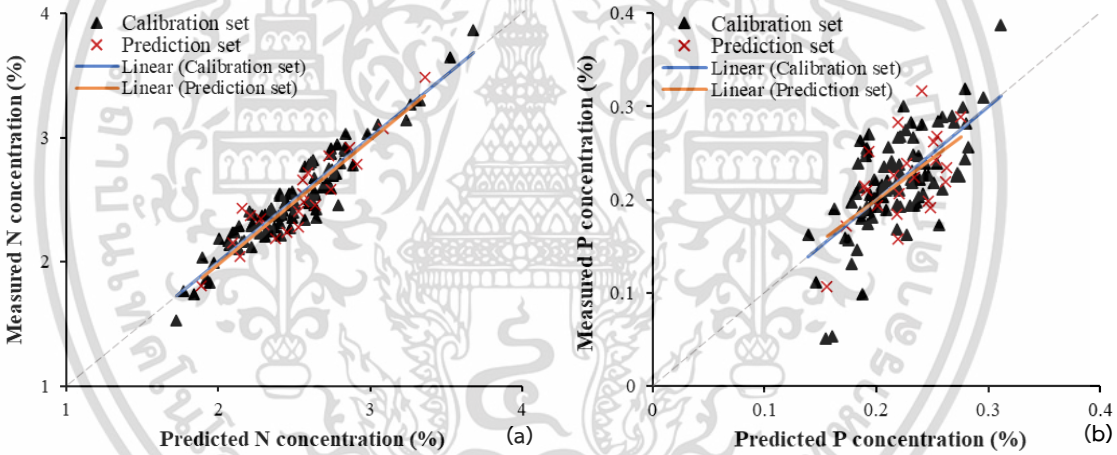
Sample[Slope t-test between models]	Parameters								
	Bias test			SEP test			Slope test		
	Bias (%)	T_b	R	SEP (%)	T_{ue}	R	T_{obs}	t value	R
N									
Fresh leaf[NS]	**-0.020**	**±0.053**	**ns**	**0.12**	**0.18**	**ns**	**0.06**	**2.08**	**ns**
Dried ground leaves powder*	0.011	±0.058	ns	0.13	0.17	ns	1.38	2.08	ns
Dried ground leaves pellet[NS]	0.045	±0.053	ns	0.13	0.16	ns	1.49	2.08	ns
K									
Fresh leaf[S]	0.043	±0.075	ns	0.17	0.22	ns	0.62	2.08	ns
Dried ground leaves powder*	**0.006**	**±0.058**	**ns**	**0.13**	**0.17**	**ns**	**0.56**	**2.08**	**ns**
Dried ground leaves pellet[NS]	-0.013	±0.058	ns	0.13	0.17	ns	0.63	2.08	ns

* R, comparison result; T_b, the bias confidence limit; SEP, standard error of prediction; T_{ue}, the unexplained error confidence limit; T_{obs}, the observed t value to check the hypothesis that slope = 1; t value, the t value obtained from the t distribution table for probability of α = 0.05; *, the highest r^2 model; NS, no significant difference between NIR predicted nutrient concentration of the highest r^2 model and itself model at 95% confidence interval; S, significant difference between NIR predicted nutrient concentration of the highest r^2 model and itself model at 95% confidence interval; ns, no significant difference at 95% confidence interval; s, significant difference at 95% confidence interval.

Table 5.3 presents the bias values of the best prediction models for N in FL, DL, and PL samples as -0.020%, 0.011%, and 0.069%, respectively. These values were lower than the bias confidence limits (T_b = ±0.053%, ±0.058%, and ±0.053%, respectively).

Regarding the K models, the bias value was lower than the bias confidence limits in all models (the biases to T_b of FL, DL, and PL samples were 0.043% to ±0.075%, 0.006% to ±0.0058%, and -0.013% to ±0.058%, respectively). These results indicate that there was no statistically significant difference between the NIR predicted and reference values.

In terms of SEP testing, the N model showed that the SEP (for FL = 0.12, DL = 0.13, and PL = 0.13) was lower than the confidence limits of unexplained error (T_{UE} for FL = 0.18, DL = 0.17, and PL = 0.16). Similarly, the SEP was lower than the T_{UE} for the K models (the SEP to T_{UE} of FL, DL, and PL were 0.17 to 0.22, 0.13 to 0.17, and 0.13 to 0.17, respectively). These results confirm that the SEP of all models can be considered acceptable.



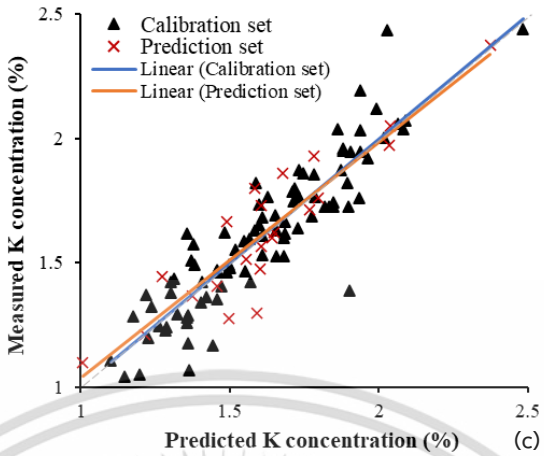


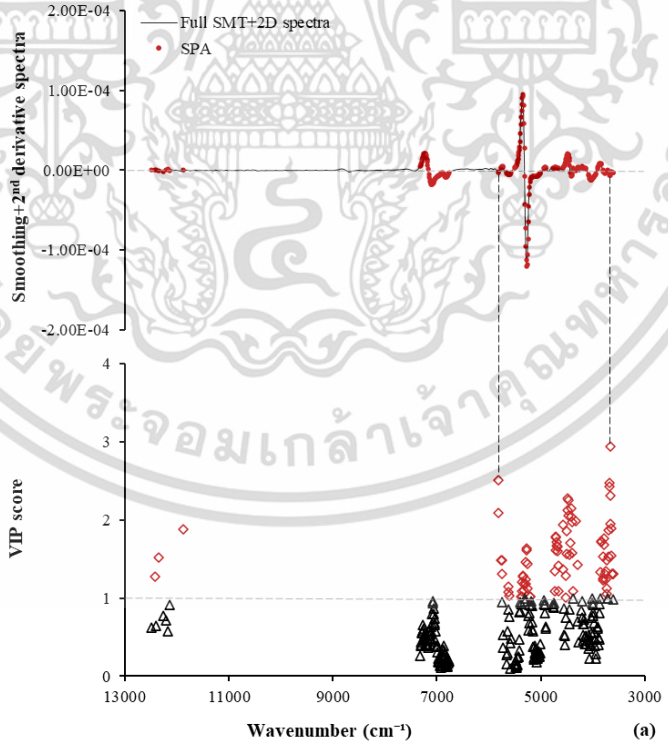
Figure 5.5 Scatter plots of measured value and predicted value from the best model to predict N content (a), P content (b) and K content (c) in durian leaves.

Furthermore, the scatter plot slopes of the prediction values and reference values for the N and K models of FL, DL, and PL were not significantly different, as indicated by t_{obs} (of N, for FL = 0.06, for DL = 1.18, and for PL = 1.49; of K, for FL = 0.62, for DL = 0.56, and for PL = 0.63) being lower than the t value (2.08). When conducting the slope hypothesis t-test for predicting macronutrients between the FL, DL, and PL models, it was observed that the N prediction models were not significantly different at the 95% confidence interval. However, the K prediction model in the FL sample differed from the other matrix models. Therefore, no sample pretreatment was required prior to NIR scanning for N estimation, but FL samples did not yield satisfactory results for K estimation. Consequently, the DL sample was deemed suitable for K prediction.

Figure 5.5 illustrates scatter plots depicting the measured values and predicted values of N content (Figure 5.5a), P content (Figure 5.5b), and K content Figure 5.5c) for the best macronutrient prediction model in durian leaf. The PLS-SPA model with SMT+2D spectra from the FL sample represented the best performance for N prediction, with r^2 = 0.852, SEP = 0.14%, RPD = 2.63, and bias = -0.020% (Table 5.2). As for P prediction, the PLS-SPA model using SMT+2D spectra and DL samples exhibited the performance, with r^2 = 0.341, SEP = 0.04%, RPD = 1.23, and bias = -0.003%. The selected model for K estimation

was the PLS-SPA model with SMT spectra from the DL sample, which yielded r^2 = 0.820, SEP = 0.13%, RPD = 2.36, and bias = 0.006%. These results indicate good prediction accuracy [35].

Figure 5.6 shows the VIP score for evaluating the N, P, and K contents. The VIP score for N prediction in fresh durian leaf was the highest in the top 5 at 5,797 cm^{-1} and 3,693 cm^{-1}, indicating that CH_2 (component of nucleic acid [36]) and CH (component of amino acid [37]) were essential in evaluating N for the PLS-SPA model. The wavelength of CH (component of sucrose [38], at 3,680 cm^{-1}) were in the top 5 VIP scores for P, meaning that CH were important to P prediction in DL sample. The NH (component of amino acids [39] that is in protein [40], at 6,805 cm^{-1}) and CH (component of amino acid [37], at 12,453 and 3,613 cm^{-1}) were key for predicting K in DL sample using NIR spectroscopy.



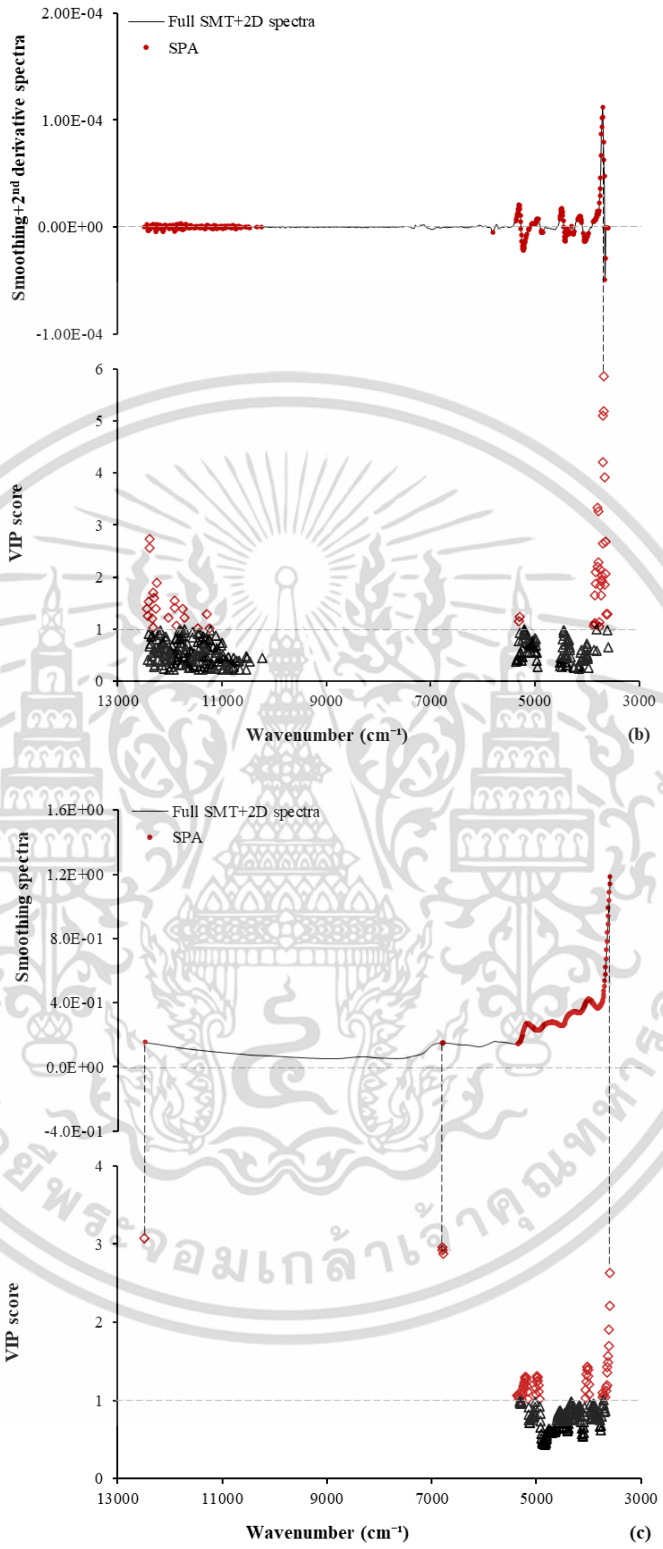


Figure 5.6 The spectrum versus VIP score for the model to predict N content (a), P content (b) and K content (c) in durian leaves.

5.4 Discussion

5.4.1 Data on the macronutrient content of durian leaf samples

The calibration set was used to create a model for macronutrient estimation. Afterward, the prediction set was used to verify the performance of the model. Both datasets of fresh leaf (FL), dried ground leaf (DL), and dried ground leaf pellet (PL) samples were from the same sample for a clear model comparison. The values of N and P in durian leaf samples covered high-appropriate-low levels, according to the standard value [6] for macronutrient concentration levels in durian leaves, and the K value covered appropriate-low levels. This was not a regression modeling problem because the K data were normally distributed.

5.4.2 NIR spectrum of durian leaf sample matrices

Although the DL and PL samples had the same particle size, the arrangement of the particles was different, which caused a baseline shift of the spectrum due to the different light scattering path lengths inside the sample [9]. However, both NIR absorption spectral structures were quite similar. After the 2^{nd} derivative spectra pretreatment, some peaks that were not observed in the raw spectra appeared because the 2^{nd} derivative pretreatment solved the overlapping problem [32]. The 2^{nd} derivative spectra showed peaks of cellulose (CH_2 [41]), proteins (amino acid and ROH [40, 42]), and starch that are related to N, P and K [10, 43-45].

5.4.3 Comparison of predicted macronutrient concentration models developed by different durian leaf sample matrices

Spectrum pretreatment and wavelength selection can improve the performance of the model because of the removal of undesirable spectral variations and noise [46]. The SPA was used to help discard noninformative variables [47] that did not make the modeling require many wavelengths, but the performance was still comparable to that of full-wavelength or short-wavelength modeling. Although the value of r^2 was slightly

decreased, the processing time was reduced due to the smaller number of wavelengths. When ISO 12099 standards were examined, the macronutrient NIR predicted, and reference nutrient were not different in the statistical significance at the 95% confidence interval. In this case, using FL samples for evaluation N in durian leaves was the recommended option since it eliminates the need for sample preparation and saves time (for one sample, the 20 leaves were scanned in 20 minutes) and using DL samples for evaluation K in durian leaves (one sample, 48 hrs for sample preparing and 1 minutes for scanning). The r^2 and RPD of N and K of durian leaf models confirmed that the models were good prediction [35]. The concentration of N has a significant influence on amino acid and protein synthesis [48], and there was a high correlation (the highest = 0.72) between NIR absorption of SMT+2D spectra and N content in fresh leaves. Similarly, the concentration of K affects protein synthesis [49], and the best correlation between NIR absorption of SMT spectra and K content in dried ground leaves was 0.58. Although P content is also important for the sucrose ratio in the leaves [50], this study was unable to estimate the P concentration due to the low correlation (<0.40) between NIR absorption of SMT+2D spectra and P content in durian leaves, and the P content range was too small. More samples, particularly those with P sample and high K concentration samples, will be needed in the future to improve the performance of the model and to cover all levels concentration range.

The use of NIR spectroscopy to evaluate macronutrient contents (N, P, and K) in leaves has been studied, but there has been no research on durian leaves. Aside from that study, the rapid measurement of N and K concentration levels in durian (CV. Monthong) leaves using an FT-NIR spectrometer is recommended [5], NIR spectroscopy can be used to identify the class of macronutrient concentrations (N and K) in durian leaves. However, further research is needed to more try in P estimation of durian leaf. As a result, users should select a model to evaluate the value or classify nutrient concentrations in durian leaves based on their intended function.

5.5 Conclusion

This work describes the application of NIR measurements to evaluate the concentrations of N and K in durian (*Durio zibethinus* Murray CV. Monthong) leaves across different sample matrices. The study highlights the following key findings: 1) FT-NIR spectroscopy is a recommended instrument for rapidly evaluating macronutrients in durian leaves using PLS models. 2) Spectral feature wavelength selection techniques, such as SPA, can aid in determining the most significant wavelengths for modeling. In particular, wavelength of CH_2 is important for N models, the wavelength of CH is important for both N and K models, and the wavelength of NH is important for the K model. 3) All three sample matrices (fresh leaf, dried ground leaf, and dried ground leaf pellet samples) can be used to develop models for N content prediction with comparable accuracy. However, for rapid preparation, fresh leaf samples are recommended as they do not require sample preparation before NIR scanning. Dried ground leaves and dried ground leaf pellet samples can also be used with good accuracy for K content prediction modeling. The prediction of P content in durian leaf using NIR spectroscopy is not successful. This study achieved good prediction models of N and K concentration by collecting samples from different stages of growth of durian trees and various locations, enabling the estimation of a wide range of nutrient contents in durian leaves. However, it is important to increase the number of samples for future predictions.

5.6 Reference

[1] McGrath, J. M., Spargo, J., and Penn, C. J. "Soil Fertility and Plant Nutrition". **Book title.**: 2014.

[2] DETURK, E. E. "Plant nutrient deficiency symptoms" **Industrial and engineering chemistry**, vol. 1941. https://doi.org/10.1021/ie50377a022.

[3] Du, Q. J., Xiao, H. J., Li, J. Q., Zhang, J. X., Zhou, L. Y., and Wang, J. Q. "Effects of different fertilization rates on growth, yield, quality and partial factor productivity of tomato under non-pressure gravity irrigation" **PLoS One**, vol. 16. 2021. pp.e0247578. https://doi.org/10.1371/journal.pone.0247578.

[4] Office of Agricultural Economics. "Durian Export Statistics" [Online]. Available: https://www.oae.go.th/view. 2023.

[5] Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "Rapid measurement of classification levels of primary macronutrients in durian (Durio zibethinus Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling" **Measurement**, vol. 203. 2022. https://doi.org/10.1016/j.measurement.2022.111975.

[6] Department of Agriculture. "Nutrient Management and Fertilizing Durian" [Online]. Available: https://www.doa.go.th/share/attachment.php?aid=2975. 2002.

[7] Phetpan, K., Udompetaikul, V., and Sirisomboon, P. "In-line near infrared spectroscopy for the prediction of moisture content in the tapioca starch drying process" **Powder Technology**, vol. 345. 2019. pp.608-615. https://doi.org/10.1016/j.powtec.2019.01.050.

[8] Quintelas, C., Braga, A., Cordeiro, A., Ferreira, E. C., Belo, I., and Páscoa, R. N. M. J. "FT-NIR spectroscopy analysis for monitoring the microbial production of 2-phenylethanol using crude glycerol as carbon source" **LWT**, vol. 155. 2022. pp.112951. https://doi.org/10.1016/j.lwt.2021.112951.

[9] Gilmer-Osborne, B., Fearn, T., Hindle, P. H., and Hindle, P. T. **Practical Nir Spectroscopy With Applications in Food and Beverage Analysis.** UK: Longman group. 1993.

[10] Wang, Y.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., and Zhang, Z.-Z. "NIR hyperspectral imaging coupled with chemometrics for nondestructive

assessment of phosphorus and potassium contents in tea leaves" **Infrared Physics & Technology**, vol. 108. 2020. https://doi.org/10.1016/j.infrared.2020.103365.

[11] Katherine L. Bolster, M. E. M., John Aber. "Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectances: a comparison of statistical methods" **Canadian Journal of Forest Research**, vol. 26(4). 1996. pp.590-600.

[12] Galvez-Sola, L., Garcia-Sanchez, F., Perez-Perez, J. G., Gimeno, V., Navarro, J. M., Moral, R., Martinez-Nicolas, J. J., and Nieves, M. "Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy" **Frontiers in Plant Science**, vol. 6. 2015. pp.571. https://doi.org/10.3389/fpls.2015.00571.

[13] Hattey, J. A., Sabbe, W. E., Baten, G. D., and Blakeney, A. B. "Nitrogen and starch analysis of cotton leaves using near infrared reflectance spectroscopy (NIRS)" **Communications in Soil Science and Plant Analysis**, vol. 25. 2008. pp.1855-1863. https://doi.org/10.1080/00103629409369158.

[14] Rebufa, C., Pany, I., and Bombarda, I. "NIR spectroscopy for the quality control of Moringa oleifera (Lam.) leaf powders: Prediction of minerals, protein and moisture contents" **Food Chemistry**, vol. 261. 2018. pp.311-321. https://doi.org/10.1016/j.foodchem.2018.04.066.

[15] Petisco, C., Garcia-Criado, B., Vazquez de Aldana, B. R., Zabalgogeazcoa, I., Mediavilla, S., and Garcia-Ciudad, A. "Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species" **Analytical and Bioanalytical Chemistry**, vol. 382. 2005. pp.458-65. https://doi.org/10.1007/s00216-004-3046-7.

[16] Rossa, Ü. B., Angelo, A. C., Nisgoski, S., Westphalen, D. J., Frizon, C. N. T., and Hoffmann-Ribani, R. "Application of the NIR Method to Determine Nutrients in Yerba Mate (Ilex paraguariensisA. St.-Hill) Leaves" **Communications in Soil Science and Plant**

Analysis, vol. 46. 2015. pp.2323-2331. https://doi.org/10.1080/00103624.2015.1081697.

[17] Gonzalez-Martin, I., Hernandez-Hierro, J. M., and Gonzalez-Cabrera, J. M. "Use of NIRS technology with a remote reflectance fibre-optic probe for predicting mineral composition (Ca, K, P, Fe, Mn, Na, Zn), protein and moisture in alfalfa" **Analytical and Bioanalytical Chemistry**, vol. 387. 2007. pp.2199-205. https://doi.org/10.1007/s00216-006-1039-4.

[18] Santoso, H., Tani, H., Wang, X., and Segah, H. "Predicting oil palm leaf nutrient contents in kalimantan, indonesia by measuring reflectance with a spectroradiometer" **International Journal of Remote Sensing**, vol. 40. 2018. pp.7581-7602. https://doi.org/10.1080/01431161.2018.1516323.

[19] Susan Ciavarella, G. D. B., and Anthony B. Blakeney. "Measuring Potassium in Plant Tissues Using near Infrared Spectroscopy" **Journal of Near Infrared Spectroscopy**, vol. 6. 1998. pp.63-66. https://doi.org/10.1255/jnirs.167.

[20] de Aldana, B. R. V., Criado, B. G., Ciudad, A. G., and Corona, M. E. P. "Estimation of mineral content in natural grasslands by near infrared reflectance spectroscopy" **Communications in Soil Science and Plant Analysis**, vol. 26. 2008. pp.1383-1396. https://doi.org/10.1080/00103629509369379.

[21] Menesatti, P., Antonucci, F., Pallottino, F., Roccuzzo, G., Allegra, M., Stagno, F., and Intrigliolo, F. "Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [Citrus sinensis (L) Osbeck cv Tarocco]" **Biosystems Engineering**, vol. 105. 2010. pp.448-454. https://doi.org/10.1016/j.biosystemseng.2010.01.003.

[22] Xiong, Y., Ohashi, S., Nakano, K., Jiang, W., Takizawa, K., Iijima, K., and Maniwara, P. "Application of the radial basis function neural networks to improve the nondestructive Vis/NIR spectrophotometric analysis of potassium in fresh lettuces"

Journal of Food Engineering, vol. 298. 2021. https://doi.org/10.1016/j.jfoodeng.2020.110417.

[23] Prananto, J. A., Minasny, B., and Weaver, T. "Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues". **Book title.**: 2020.

[24] Guo, J., Huang, H., He, X., Cai, J., Zeng, Z., Ma, C., Lü, E., Shen, Q., and Liu, Y. "Improving the detection accuracy of the nitrogen content of fresh tea leaves by combining FT-NIR with moisture removal method" **Food Chemistry**, vol. 405. 2023. https://doi.org/10.1016/j.foodchem.2022.134905.

[25] Shi, J., Wang, Y., Li, Z., Huang, X., Shen, T., and Zou, X. "Simultaneous and nondestructive diagnostics of nitrogen/magnesium/potassium-deficient cucumber leaf based on chlorophyll density distribution features" **Biosystems Engineering**, vol. 212. 2021. pp.458-467. https://doi.org/10.1016/j.biosystemseng.2021.11.001.

[26] Yao, L., Shi, X., Pan, T., and Chen, J. "Wavelength Selection Method Based on Absorbance Value Optimization to Near-Infrared Spectroscopic Analysis" **Frontiers in Physics**, vol. 9. 2021. https://doi.org/10.3389/fphy.2021.663573.

[27] He, H.-J., Chen, Y., Li, G., Wang, Y., Ou, X., and Guo, J. "Hyperspectral imaging combined with chemometrics for rapid detection of talcum powder adulterated in wheat flour" **Food Control**, vol. 144. 2023. https://doi.org/10.1016/j.foodcont.2022.109378.

[28] Maraphum, K., Saengprachatanarug, K., Wongpichet, S., Phuphuphud, A., and Posom, J. "Achieving robustness across different ages and cultivars for an NIRS-PLSR model of fresh cassava root starch and dry matter content" **Computers and Electronics in Agriculture**, vol. 196. 2022. https://doi.org/10.1016/j.compag.2022.106872.

[29] Udompetaikul, V., Phetpan, K., and Sirisomboon, P. "Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an

elevator conveyor" **Measurement**, vol. 167. 2021. https://doi.org/10.1016/j.measurement.2020.107898.

[30] Rutledge, D. N., Roger, J.-M., and Lesnoff, M. "Different Methods for Determining the Dimensionality of Multivariate Models" **Frontiers in Analytical Science**, vol. 1. 2021. https://doi.org/10.3389/frans.2021.754447.

[31] Mohd Asaari, M. S., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., and Scheunders, P. "Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform" **ISPRS Journal of Photogrammetry and Remote Sensing**, vol. 138. 2018. pp.121-138. https://doi.org/10.1016/j.isprsjprs.2018.02.003.

[32] Williams, P. and Norris, K. **Near-Infrared Technology: In the Agricultural and Food Industries.** American Association of Cereal Chemists. 1987.

[33] Pochanagone, S. and Rittiron, R. "Preliminary Study on the Determination of ppm-Level Concentration of Histamine in Tuna Fish Using a Dry Extract System for Infrared Coupled with Near-Infrared Spectroscopy" **ACS Omega**, vol. 4. 2019. pp.19164-19171. https://doi.org/10.1021/acsomega.9b02438.

[34] Ng, W., Minasny, B., Malone, B. P., Sarathjith, M. C., and Das, B. S. "Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling" **Computers and Electronics in Agriculture**, vol. 158. 2019. pp.201-210. https://doi.org/10.1016/j.compag.2019.02.003.

[35] Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., and Mataix-Beneyto, J. "Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils" **Soil Biology and Biochemistry**, vol. 40. 2008. pp.1923-1930. https://doi.org/10.1016/j.soilbio.2008.04.003.

[36] Holde, K., and Zlatanova. **The Evolution of Molecular Biology**. Elsevier Inc. 2018.

[37] Makishima, A. **Biochemistry for Materials Science.** Elsevier Inc. 2019.

[38] Plaza-Diaz, J., and Gil, A. **Encyclopedia of Food and Health**. Elsevier Inc. 2016.

[39] Maloy, S. **Brenner's Encyclopedia of Genetics.** Life Sciences. 2013.

[40] Kumar, M., Tomar, M., Punia, S., Dhakane-Lad, J., Dhumal, S., Changan, S., Senapathy, M., Berwal, M. K., Sampathrajan, V., Sayed, A. A. S., Chandran, D., Pandiselvam, R., Rais, N., Mahato, D. K., Udikeri, S. S., Satankar, V., Anitha, T., Reetu, Radha, Singh, S., Amarowicz, R., and Kennedy, J. F. "Plant-based proteins and their multifaceted industrial applications" **Lwt**, vol. 154. 2022. https://doi.org/10.1016/j.lwt.2021.112620.

[41] M, A., and Deepu, T. **Biopolymer Composites in Electronics.** Elsevier Ltd. 2017.

[42] Makishima, A. **Biochemistry for Materials Science: Catalysts, Complexes and Proteins.** Elsevier. 2019.

[43] Min, M., Lee, W. S., Kim, Y. H., and Bucklin, R. A. "Nondestructive Detection of Nitrogen in Chinese Cabbage Leaves UsingVIS–NIR Spectroscopy" **HortScience**, vol. 41. 2006. pp.162-166. https://doi.org/10.21273/HORTSCI.41.1.162.

[44] Jin, X., Wang, L., Zheng, W., Zhang, X., Liu, L., Li, S., Rao, Y., and Xuan, J. "Predicting the nutrition deficiency of fresh pear leaves with a miniature near-infrared spectrometer in the laboratory" **Measurement**, vol. 188. 2022. pp.110553. https://doi.org/10.1016/j.measurement.2021.110553.

[45] Ciavarella, S. and Batten, G. D. "Measuring potassium in plant tissues using near Infrared spectroscopy" **Journal of Near Infrared Spectroscopy** vol. 6. 1998. https://doi.org/10.1255/jnirs.167.

[46] Krepper, G., Romeo, F., Fernandes, D. D. S., Diniz, P., de Araujo, M. C. U., Di Nezio, M. S., Pistonesi, M. F., and Centurion, M. E. "Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS)" **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, vol. 189. 2018. pp.300-306. https://doi.org/10.1016/j.saa.2017.08.046.

[47] de Araújo Gomes, A., Galvão, R. K. H., de Araújo, M. C. U., Véras, G., and da Silva, E. C. "The successive projections algorithm for interval selection in PLS" **Microchemical Journal**, vol. 110. 2013. pp.202-208. https://doi.org/10.1016/j.microc.2013.03.015.

[48] Perchlik, M., and Tegeder, M. " Leaf Amino Acid Supply Affects Photosynthetic and Plant Nitrogen Use Efficiency under Nitrogen Stress" **Plant Physiology Preview**, vol.178(1), 2018. pp.174-188. https://doi.org/10.1104/pp.18.00597.

[49] Xu, X., Du, X., Wang, F., Sha, J., Chen, Q., Tian, G., Zhu, Z., Ge, S., and Jiang, Y. "Effects of Potassium Levels on Plant Growth, Accumulation and Distribution of Carbon, and Nitrate Metabolism in Apple Dwarf Rootstock Seedlings" **frontiers in plant science**, vol. 2020. https://doi.org/10.3389/fpls.2020.00904.

[50] Spencer, C. F. a. C. "The relationship between phosphate status and photosynthesis in leaves" **Planta**, vol. 167. 1986. pp.369-375.

Chapter 6

The near infrared spectroscopy model lowest limit of quantification for nutrient in durian (CV Monthong) leaf evaluation

NIR spectroscopy is a well-known method to obtain quick assessments. The Fourier transform near-infrared (FT-NIR) spectrometer can rapidly determine the macronutrient (nitrogen; N, phosphorus; P, and potassium; K) content of durian leaves (fresh leaf; FL, dried ground leaves; DL, and dried ground leaves pellet; PL) by the model. The NIR model's proven lowest limit of quantification (LOQ) is worth to study to present the lowest monitoring limit. The goal of the study was to evaluate for the LOQ of the partial least squares (PLS) model for the macronutrients evaluation in durian leaves which model developed by full-spectra (12,500-3,600 cm^{-1}), short-spectra (7,500-6,000 cm^{-1} + 6,600-4,000 cm^{-1} for FL and 6,000-4,000 cm^{-1} for DL and PL) and selected wavelength using a successive projections algorithm (SPA) and were spectra pre-treat by smoothing, smoothing+standard normal variate and smoothing+2nd derivative method. The LOQ of N model using full-spectra, short-spectra and SPA were 1.52%, 1.30% and 1.22% for FL, 0.88%, 1.21% and 1.27% for DL, and 1.64%, 1.17% and 1.30% for PL, respectively, while the minimum of N data was 1.53%. The LOQ of P model using full-spectra, short-spectra and SPA were 0.48%, 0.45% and 0.43% for FL, 0.40%, 0.33% and 0.39% for DL, and 0.36%, 0.38% and 0.41% for PL, respectively, while the minimum of P data was 0.05%. The LOQ of K model using full-spectra, short-spectra and SPA were 1.15%, 1.88% and 1.71% for FL, 1.44%, 1.24% and 1.29% for DL, and 1.27%, 1.28% and 1.33% for PL, respectively, while the minimum of K data was 1.04%. This results showed a sensitivity of NIR models for

macronutrient assessment which the LOQ of N model was lower than the minimum content for modelling content.

6.1 Introduction

Near infrared (NIR) spectroscopy is a rapid and non-destructive analysis which is widely used to qualitative and quantitative testing in pharmacy, agriculture, food, and chemical industry [1]. There are several published research on the use of NIR to predict nutrient in leaf; apple [2], tomato [3, 4], olive [5], rubber [6], orange [7], tea [8], moringa oleifera [9] and durian [10]. Our study is also successful. Chapter 5 shows that the FT-NIR spectrometer can assess macronutrient in durian (CV Monthong) leaf when it combined with PLS model for fresh leaf (FL), dried ground leaves (DL) and dried ground leaves pellet (PL) sample. The model regarding every sample matrix using smoothing $+2^{nd}$ derivative spectra was acceptable, and the prediction was considered good to excellent. The nitrogen (N) predicting model showed the r^2 of FL = 0.981-0.983, DL = 0.957-0.988, and PL = 0.971-0.992. The phosphorus (P) predicting model showed the r^2 of FL = 0.974-0.986, DL = 0.965-0.984, and PL = 0.972-0.988. The potassium (K) predicting model showed the r^2 of FL = 0.974-0.992, DL = 0.954-0.984, and PL = 0.980-0.991.

The method validation is necessary to make sure that the NIR protocol developed, and reference method is reliable [11]. The analytical performance characteristics (parameters) were accuracy, precision, specificity, limit of detection (LOD), limit of quantitation (LOQ), linearity, range, and robustness [12]. The method validation is not required to complete all parameters. The monitored parameter depends on the type of analytical method. The NIR spectroscopy method often validates precision (repeatability and reproducibility) and linearity (R^2). However, we can verify the LOQ for evaluation to the smallest or the lowest concentration which the used analyzer can accept with accuracy, precision, and uncertainty [11], which indicates sensitivity of model.

This study focuses on the LOQ of PLS regression model for macronutrient assessment in durian (CV Monthong) leaf using near infrared spectroscopy which modeling by the leaf sample matrices (fresh leaf, dried ground leaves, and dried ground leaf pellet) and wavelength selection (full wavelength, short wavelength, and selected wavelength by successive projections algorithm (SPA)). The goals of this study were as follows: 1) to validate calibration model by the LOQ. 2) to compare the LOQ from difference sample and wavelength selection model. If the LOQ is lower than minimum value for modeling, it can confirm the lowest concentration that model can detect.

6.2 Materials and Methods

6.2.1 Durian leaf samples

The fresh leaves of durian (Durio zibethinus Murray CV. Monthong) were collected from Rayong, Chanthaburi, and Trad in Thailand from December 2018 to November 2020. A sample of 20 fresh leaves was collected from one durian tree. After collecting, the 20 fresh leaf samples were washed with 0.1 N HCl acid and rinsed 3 times with distilled water. Subsequently, samples were dried at 70°C for 48 h, crushed, and sieved through 40 mesh (0.42 mm) perforated screen that obtain a dried ground leaves sample. The 0.15±0.01g dried ground leaf sample was placed in a stainless block (diameter 1.40 cm) and pressurized at 60 MPa to obtain a dried ground leaves pellet sample. A total of 89 samples were used for calibration modeling, and 22 samples were used for model validation.

6.2.2 Near infrared scanning

The fresh leaves sample (FL) was scanned in two positions (proximal and distal side) while covering with an aluminum plate. The dried ground leaf sample (DL) was scanned with by using a quartz cup (4.3 cm diameter, 7.0 cm in height). The dried ground leaf pellet sample (PL) was scanned while covering with Spectralon reference material and covered with an aluminum can. The used NIR spectrometer is FT (Fourier

transformed)-NIR spectrometer (Bruker Ltd., Germany) which scanned at a wavenumber of 12,500–4,000 cm^{-1}, an average of 32 scans and a resolution of 32 cm^{-1}.

6.2.3 Sample analysis

The dried ground leaf samples were analyzed by TruMac CNS (Leco, USA) for nitrogen (N) and ICP-OES (Perkin Elmer, USA) for phosphorus (P) and potassium (K). Table 6.1 shown macronutrient data of sample which was used modeling.

Table 6.1 Statistical characteristics of macronutrient data.

Nutrients	Datasets	N	Max (%)	Mean (%)	Min (%)	SD (%)
N	Calibration	89	3.87	2.50	1.53	0.39
	Prediction	22	3.48	2.51	1.81	0.37
P	Calibration	89	0.39	0.22	0.05	0.05
	Prediction	22	0.32	0.22	0.11	0.05
K	Calibration	89	2.44	1.61	1.04	0.30
	Prediction	22	2.37	1.62	1.10	0.30

*N: number of samples, SD: the standard deviation.

6.2.4 Modeling

This work used regression model (from chapter 5) by partial least squares (PLS) regression with smoothing $+2^{nd}$ derivative spectra. The wavelengths used for modeling were the full wavelength (12,500-3,600 cm^{-1}), short wavelength (7,500-6,600 cm^{-1} + 6,000-4,000 cm^{-1} for FL and 6,000-4,000 cm^{-1} for DL and PL), and selected wavelength by successive projections algorithm (SPA).

6.2.5 The limit of quantification

The limit of quantitation (LOQ) or sometimes also referred as quantification limit, (QL) is the lowest possible concentration of analyte that can be reliably measured by the method [13]. There are several ways to find LOQ: based on visual evaluation, based on

signal-to-noise, and based on the standard deviation of the response on the slope [14]. This study focuses on the LOQ based on the standard deviation of the response on the slope. The LOQ can be calculated as this equation:

$$LOQ = \frac{10\sigma}{S} \quad (6.1)$$

where σ is the residual standard deviation of reference value and predicted value.

S is the slope of the regression line.

Figure 6.1 shows the step of the lowest limit of quantification (LOQ) determining.

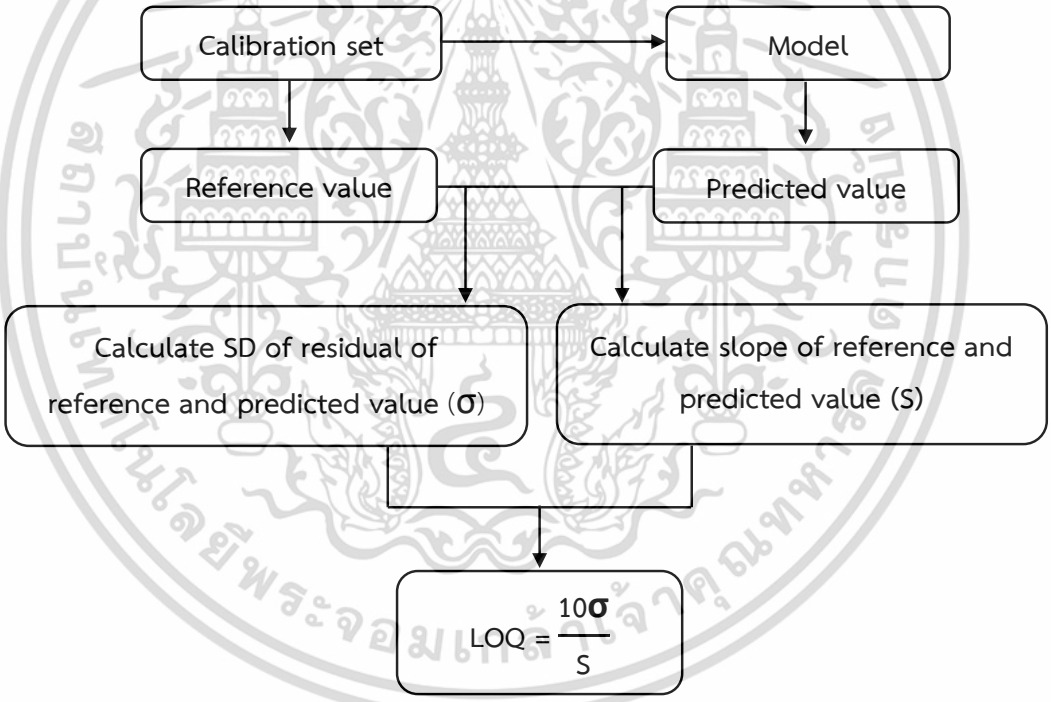


Figure 6.1 Step of determining the lowest limit of quantification (LOQ).

6.3 Results

The PLS regression models from chapter 5 were used to determine the LOQ to validate model. Table 6.2 shows the LOQ of macronutrients content assessment model in durian leaf.

Table 6.2 The LOQ result of PLS models for macronutrients content assessment in durian leaf.

Nutrient	Sample	Spectra	pretreatment	nLVs	nWL	R^2	r^2	SD	slope	LOQ
N	FL	Full	SMT	12	1154	0.845	0.815	0.15	1.00	1.52
		Short	SMT	15	365	0.885	0.851	0.13	1.00	1.30
		SPA	SMT+2D	15	286	0.9	0.852	0.12	1.00	1.22
	DL	Full	SMT	14	1154	0.948	0.912	0.09	1.00	0.88
		Short	SMT+SNV	14	182	0.902	0.869	0.12	1.00	1.21
		SPA	SMT	11	235	0.891	0.879	0.13	1.00	1.27
	PL	Full	SMT	8	1154	0.838	0.786	0.16	1.00	1.64
		Short	SMT+SNV	12	182	0.909	0.851	0.12	1.00	1.17
		SPA	SMT	11	235	0.887	0.880	0.13	1.00	1.30
P	FL	Full	SMT	3	1154	0.154	0.232	0.05	1.00	0.48
		Short	SMT+2D	4	378	0.283	0.235	0.04	1.00	0.45
		SPA	SMT+2D	5	372	0.327	0.323	0.04	1.00	0.43
	DL	Full	SMT+2D	3	1154	0.411	0.276	0.04	1.00	0.40
		Short	SMT+2D	6	260	0.557	0.100	0.03	1.00	0.33
		SPA	SMT+2D	3	361	0.447	0.341	0.04	1.00	0.39
	PL	Full	SMT+SNV	9	1154	0.523	0.258	0.04	1.00	0.36
		Short	SMT+2D	5	260	0.485	0.362	0.04	1.00	0.38
		SPA	SMT+2D	6	150	0.393	0.220	0.04	1.00	0.41
K	FL	Full	SMT+2D	8	1154	0.849	0.678	0.12	1.00	1.15
		Short	SMT+SNV	10	365	0.602	0.524	0.19	1.00	1.88
		SPA	SMT+2D	14	45	0.67	0.657	0.17	1.00	1.71
	DL	Full	SMT	10	1154	0.765	0.758	0.14	1.00	1.44
		Short	SMT+SNV	18	182	0.825	0.834	0.12	1.00	1.24
		SPA	SMT	14	234	0.811	0.820	0.13	1.00	1.29
	PL	Full	SMT	13	1154	0.817	0.836	0.13	1.00	1.27
		Short	SMT+SNV	19	182	0.814	0.747	0.13	1.00	1.28
		SPA	SMT+SNV	11	971	0.800	0.808	0.13	1.00	1.33

*nLVs, number of latent variables (PLS factors); nWL, number of wavelengths; R^2, coefficient of determination of calibration set; r^2, coefficient of determination of prediction set; SD, standard deviation; LOQ, lowest limit of detection quantification; SPA, successive projections algorithm; SMT, smoothing; SMT+SNV, smoothing+standard normal variate; SMT+2D, smoothing+2nd derivative

6.4 Discussion

The LOQ of all N content models (except for a model for PL samples using SMT spectra and the full wavelength) was lower than the minimum N concentration used for modeling. This means that the models can detect N concentrations in durian leaves from approximately 1.53% onwards or based on the LOQ value of the respective model. The LOQ of the P content model was higher than the minimum P concentration used for modeling (above 0.28-0.43%). This indicates that the models have limitations for evaluating P content in durian leaves. Predicted samples should have a P concentration higher than the model's LOQ. Similarly, the LOQ of the K content model was also higher than the lowest concentration utilized for modeling (above 0.11-0.84%). Consequently, the models have limitations in evaluating K content in durian leaves. The estimated sample should contain more K than the model's LOQ. It was observed that as the R^2 value of a model increased, the LOQ decreased, as a more accurate model leads to more precise predictions and therefore the sensitivity of the model. Consequently, the lowest possible detection limit is also reduced.

One limitation of using NIR for nutrient content prediction is that the modeling data must encompass all future data comprehensively. Therefore, collecting a large number of samples is necessary for this method. Naturally, we cannot control the nutrient content of durian leaves as needed by fertilizing [15]. Nutrients are absorbed by plants from the soil and stored in various parts, such as leaves, trunks, flowers, and fruits. However, there is one way to create a sample with high nutrient content in the leaf, which is through foliar spraying. Foliar plant spray involves directly applying fertilizer to the leaves of a plant. On the other hand, there is no method to deliberately create a sample with low nutrient content in the leaf, except by refraining from using fertilizer. Consequently, collecting samples with low nutrient content poses greater difficulty. Therefore, an accurate model can help overcome the challenge of collecting leaf samples with low nutrient content, as the LOQ of the NIR model demonstrates sensitivity lower than the minimum content of a real sample.

6.5 Conclusion

The lowest limit of quantification of NIR models for evaluating N concentration was lower than the minimum reference value used for modeling, indicating high sensitivity of NIR models for assessing N. However, the lowest quantification limit of models for estimating P and K was higher than the minimal reference value. Consequently, future improvement of the model accuracy will improve its sensitivity too.

6.6 Reference

[1] Wu, Z., Du, M., Sui, C., Lin, Z., Shi, X., and Qiao, Y. "Feasibility Analysis of Lower Limit of Quantification of NIR for Solvent in Different Hydrogen Bonds Environment Using Multivariate Calibrations," presented at the 2012 International Conference on Biomedical Engineering and Biotechnology, 2012.

[2] Azadnia, R., Rajabipour, A., Jamshidi, B., and Omid, M. "New approach for rapid estimation of leaf nitrogen, phosphorus, and potassium contents in apple-trees using Vis/NIR spectroscopy based on wavelength selection coupled with machine learning" **Computers and Electronics in Agriculture**, vol. 207. 2023. https://doi.org/10.1016/j.compag.2023.107746.

[3] Lequeue, G., Draye, X., and Baeten, V. "Determination by near infrared microscopy of the nitrogen and carbon content of tomato (Solanum lycopersicum L.) leaf powder" **Sci Rep**, vol. 6. 2016. pp.33183. https://doi.org/10.1038/srep33183.

[4] Ulissi, V., Antonucci, F., Benincasa, P., Farneselli, M., Tosti, G., Guiducci, M., Tei, F., Costa, C., Pallottino, F., Pari, L., and Menesatti, P. "Nitrogen concentration estimation in tomato leaves by VIS-NIR non-destructive spectroscopy" **Sensors (Basel)**, vol. 11. 2011. pp.6411-24. https://doi.org/10.3390/s110606411.

[5] Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., and Yermiyahu, U. "Estimating olive leaf nitrogen concentration using visible

and near-infrared spectral reflectance" **Biosystems Engineering**, vol. 114. 2013. pp.426-434. https://doi.org/10.1016/j.biosystemseng.2012.09.005.

[6] Guo, P.-T., Li, M.-F., Luo, W., and Cha, Z.-Z. "Estimation of foliar nitrogen of rubber trees using hyperspectral reflectance with feature bands" **Infrared Physics & Technology**, vol. 102. 2019. pp.103021. https://doi.org/10.1016/j.infrared.2019.103021.

[7] Menesatti, P., Antonucci, F., Pallottino, F., Roccuzzo, G., Allegra, M., Stagno, F., and Intrigliolo, F. "Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [Citrus sinensis (L) Osbeck cv Tarocco]" **Biosystems Engineering**, vol. 105. 2010. pp.448-454. https://doi.org/10.1016/j.biosystemseng.2010.01.003.

[8] Wang, Y.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., and Zhang, Z.-Z. "NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves" **Infrared Physics & Technology**, vol. 108. 2020. https://doi.org/10.1016/j.infrared.2020.103365.

[9] Rébufa, C., Pany, I., and Bombarda, I. "NIR spectroscopy for the quality control of Moringa oleifera (Lam.) leaf powders: Prediction of minerals, protein and moisture contents" **Food Chemistry**, vol. 261. 2018. pp.311-321. https://doi.org/10.1016/j.foodchem.2018.04.066.

[10] Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "Rapid measurement of classification levels of primary macronutrients in durian (Durio zibethinus Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling" **Measurement**, vol. 203. 2022. https://doi.org/10.1016/j.measurement.2022.111975.

[11] Konieczka, P., "Validation and Regulatory Issues for Sample Preparation," **Comprehensive Sampling and Sample Preparation**, 2012, pp. 699-711.

[12] duangteerapeecha, S., "Analytical Method Validation," Department of Medical Sciences, Ministry of Public Health, 2018.

[13] Lösungsfabrik. "ICH Q2(R1) method validation limit of detection limit of quantification" [Online]. Available: https://mpl.loesungsfabrik.de/en/english-blog/method-validation/limit-of-detection-quantifcation. 2023.

[14] ICH harmonised tripartite guideline, "validation of analytical procedures: text and methodology Q2(R1)," 2005.

[15] Poovarodom, S. and Phanchindawan, N. "Effects of Chloride and Sulfate in Various N and K Fertilizers on Soil Chemical Properties and Nutrient Concentrations in Durian Leaf and Fruit" **Acta Horticulturae**, vol. 271. 2006. pp.191-198. https://doi.org/10.17660/ActaHortic.2006.721.25.

Chapter 7

The evaluation of nitrogen in durian (CV Monthong) leaf using terahertz (THz) spectroscopy

Terahertz (THz) spectroscopy is a nondestructive technique that the study of the interaction between matter and electromagnetic radiation with frequencies in the terahertz range (typically between 0.1 and 10 THz). The durian (*Durio zibethinus* Murray CV. Monthong) leaf tablet was scanned by Terahertz time-domain (THz-TD) spectrometer with a bandwidth of 0.1-4.0 THz. Partial least squares (PLS) regression models with full wavelength were created to nitrogen (N) evaluate in the sample. This result shows that relationship of THz spectra and N content was not found.

7.1 Introduction

Nitrogen (N) is an essential element for plant growth and is a crucial component of leaf tissue. It plays a vital role in various physiological processes, including photosynthesis, protein synthesis, and chlorophyll production. In plant leaves, N is primarily present in the form of proteins and chlorophyll [1]. When N is deficient, it can lead to stunted growth, yellowing of leaves, and reduced overall plant health [2]. In durian cultivation, the proper amount of nitrogen is required for the tree to grow up and to be ready for durian fruit production.

Terahertz spectroscopy is a technique of studying and analyzing materials using terahertz (THz) radiation. Terahertz radiation lies in the electromagnetic spectrum between the microwave and infrared regions, with frequencies typically ranging from 0.1 to 10 THz (10^{12} Hz) [3]. It is safe for investigating biological samples, including tissues and cells, since

it is low energy radiation that can penetrate dry, nonmetallic, nonpolar materials but is not powerful enough to ionize an atom or a molecule [4]. Terahertz technology is currently the most powerful for characterization of complicated dielectric materials [5]. As a result, THz has applications in non-destructive testing (NDT), food industry, communication, medical and biomedical, agriculture, and safety and security [6] fields. Terahertz Time-Domain Spectroscopy (THz-TDS) is one method commonly used in THz spectroscopy that involves ultrashort pulses of terahertz radiation generated using femtosecond lasers. These pulses are then directed onto the sample, and the response of the sample is measured as a function of time [7]. By analyzing the time-domain signal, information about the dielectric properties, refractive index, and absorption features of the sample can be obtained.

THz spectroscopy is being used in study to identify features in leaves such as electromagnetic properties [8], water content [9, 10], chlorophyll [11] and nitrogen [12]. From research on terahertz vibrational modes of sodium magnesium chlorophyllin and chlorophyll in plant leaves, it was found that chlorophyll a were observed at 1.45, 1.73, and 1.86 THz [11]. From research on a nondestructive model for the detection of the nitrogen content of tomato, it was found that THz spectroscopy can measure the nitrogen content of tomato leaves (the peak showed at around 0.7 THz.) [12]. Although no research has used THz spectroscopy to assess nitrogen in durian leaves, there is research that can measure nitrogen in the leaves, and nitrogen is necessary for chlorophyll. This research aimed to study the feasibility of estimating the nitrogen content in durian leaves using Terahertz Time-Domain Spectroscopy.

7.2 Materials and Methods

7.2.1 Durian leaf samples

From December 2018 to November 2020, fresh durian (*Durio zibethinus* Murray CV. Monthong) leaves were collected in Rayong, Chanthaburi, and Trad in Thailand. Twenty fresh leaves around one durian tree was collected to represent one sample (dark green leaf was chosen) [13]. The fresh leaf samples were washed with 0.1 N HCl acid, rinsed 3 times with distilled water, dried for 48 hours at 70°C, crushed, and sieved through a 40 mesh (0.42 mm) that obtain a leaf powder. After that a 0.13 ± 0.01g leaf powder was put in a 1.40 cm diameter stainless block measurement (Figure 7.1a) and pressurized at 60 MPa measurement (Figure 7.1b) to obtain a tablet sample (Figure 7.1c) for THz measurement. Total samples were 111 samples.

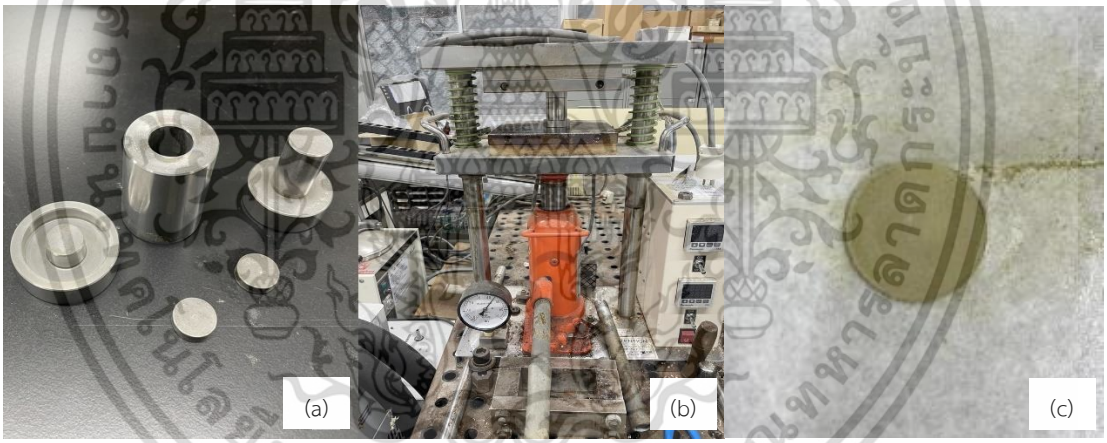


Figure 7.1 The stainless block (a), pressure machine (b) and sample (c) for THz measurement.

7.2.2 Terahertz time-domain (THz-TD) measurement

The terahertz (THz) spectra were measured by a Tera Prospector (Nippo, Precision Co., Ltd., Japan) with horizontally polarized beam, a bandwidth of 0.1-4.0 THz, an average of 100 scans and the spectral resolution was 0.02 THz. The air was used as the reference signals both before and after the sample measurement. The measurement was controlled

relative humidity in THz optical system (Figure 7.2a) at 0.0-0.1% and controlled under room conditions at 26°C. Before measuring, tablet samples were placed in an almost-closed acrylic box 24 hrs to balance the ambient humidity. One sample was measured in 3 replicates and averaged. One time for THz measurement, it can put 5 samples on a holder plate (Figure 7.2b).

The THzTDS program was processed to obtain the phase difference between the reference and measured samples (ϕ, °) and the ratio of the amplitude in the frequency domain of the measured samples to the reference (R). After that the ϕ and R value were transferred to absorption coefficient (α) by Matlab from formula [14];

$$\alpha = -\frac{2}{L} \ln \left[R \frac{(n+1)^2}{4n} \right] \quad (7.1)$$

$$n = -\frac{\phi c}{2\pi v L} + 1 \quad (7.2)$$

where n is refractive index value

c is the speed of light (3×10^8 m s^{-1})

v is the frequency (THz)

L is the thickness of the samples (the durian leaf tablet samples had a thickness of $0.96 \pm 0.03 \times 10^{-3}$ m)

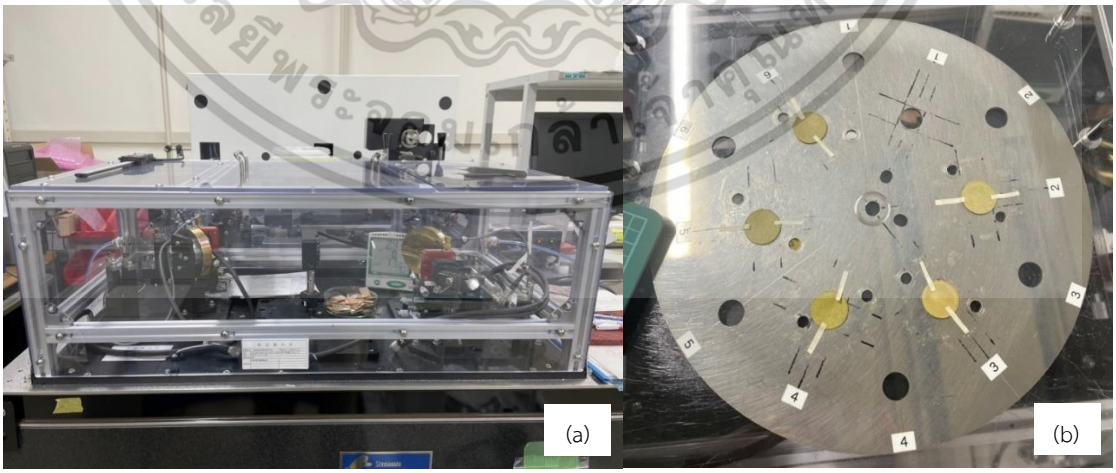


Figure 7.2 The THz spectrometer (a) and sample holder plate (b).

7.2.3 Nitrogen analysis

TruMac CNS (Leco, USA) was used to determine the nitrogen (N) concentration with 0.10 g leaf powder. Before measurement, the sample was dried at 70°C for 48 hours to make sure that the sample was completely dried.

7.2.4 Model and performance

In this study, the samples were ordered in ascending order and speared 80:20 to calibration set (89 samples) and prediction set (22 samples), which calibration data covered prediction data. The spectrum was nontreatment and pretreatment before modeling which included 1) raw spectra, 2) smoothing (SMT) spectra, and 3) smoothing + standard normal variate (SMT+SNV) spectra. The calibration samples were used to create the leave-one-out cross-validation model by the partial least squares (PLS) regression with full wavelength (0.01-0.40 THz). The PLS regression is linear multivariate model by the relationships between two data matrices [15], X (absorbance) and Y (nitrogen). It is finding latent variables (LVs) which are linear combinations of the X and Y data that explain maximizes the covariance between the X and Y data [16].

The models were validated to estimate model performance, and the following indices were calculated: the coefficient of determination of the cross-validation set (R^2), the standard deviation of error of cross-validation (SECV), the coefficient of determination of the prediction set (r^2), the standard deviation of error of prediction (SEP), bias, and the ratio of prediction to deviation (RPD), which were calculated by the following formulas:

$$R^2, r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5.1)$$

$$SECV, SEP = \sqrt{\frac{\sum \left(\left(y_i - \hat{y}_{\bar{i}} \right) - \frac{\sum (y_i - \hat{y}_i)}{n} \right)^2}{n-1}} \quad (5.2)$$

$$Bias = \frac{\sum y_i - \hat{y}_i}{n} \quad (5.3)$$

$$RPD = \frac{SD_y}{SEP} \quad (5.4)$$

where y_i is the reference value

$\hat{y}_i$ is the predicted value

$\bar{y}$ is the average of the reference value

n is the number of samples.

SD_y is the standard deviation of the reference value of the prediction set. Following the selection of the models, the r^2 and RPD values were used to assess the model's performance [17].

7.3 Results

Figure 7.3 shows THz spectrum which nontreatment (Figure 7.3a) and pretreatment by SMT (Figure 7.3b), and SMT+SNV (Figure 7.3c). When the relationship between absorption coefficient (**α**) and N content was investigated, the correlation of raw, SMT, and SMT+SNV spectra with N content was lower than 0.31 (Table 7.1).

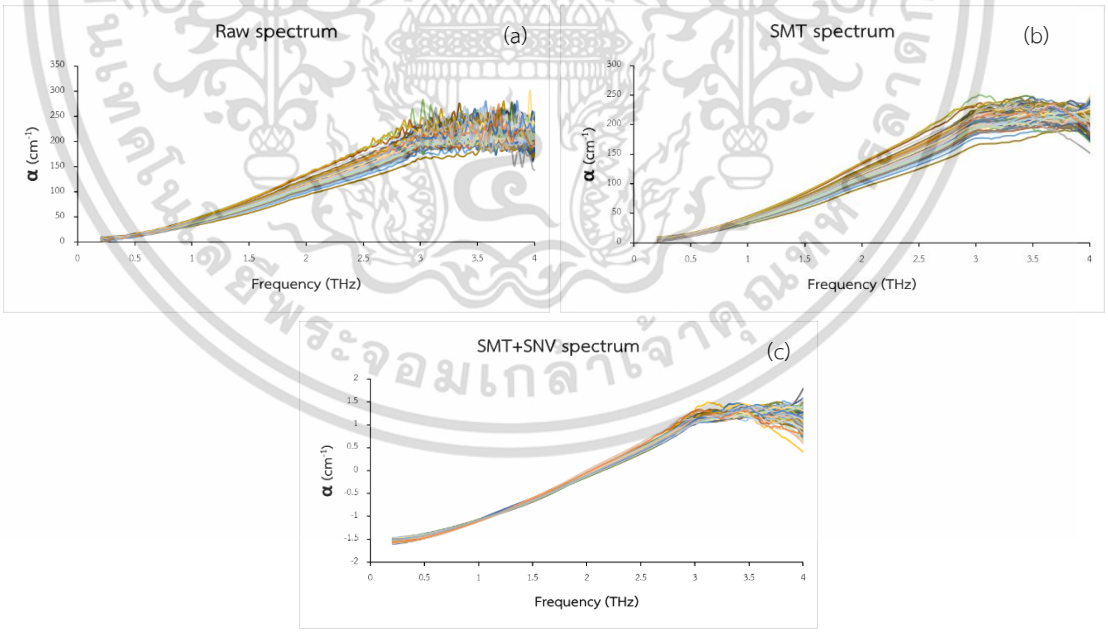


Figure 7.3 The raw (a), SMT (b), and SMT+SVN (c) spectrum.

To develop a model for predicting N in durian leaves using Terahertz time-domain spectroscopy (THZ-TDs) and PLS regression, the number of LVs (nLVs) varied from 1 to 20 to compare the best model. Table 7.2 shows the best result of model's performance.

Table 7.1 The correlation between THz spectroscopy and N content.

	Raw spectra		SMT spectra		SMT+SNV spectra	
	f (THz)	cor	f (THz)	cor	f (THz)	cor
top 10 highest	3.16	0.31	2.72	0.27	2.72	0.26
	2.72	0.29	2.82	0.27	2.70	0.25
	3.09	0.28	2.75	0.27	2.75	0.25
	2.82	0.28	2.79	0.27	2.68	0.25
	2.68	0.27	2.70	0.27	2.79	0.25
	3.18	0.27	2.68	0.27	2.82	0.24
	2.75	0.27	2.77	0.27	2.77	0.24
	2.77	0.27	2.66	0.26	2.66	0.24
	2.59	0.27	2.84	0.26	2.63	0.24
	2.63	0.26	3.48	0.26	2.61	0.23
From research	0.71	0.16	0.71	0.16	0.71	-0.06
[11, 12]	1.45	0.18	1.45	0.17	1.45	-0.09
	1.73	0.19	1.73	0.19	1.73	-0.05
	1.86	0.17	1.86	0.18	1.86	-0.07

*SMT, smoothing; SNV, Vector Normalization; f, frequency (THz); cor, correlation.

The PLS model with raw spectrum was obtained R^2 = 0.146, SECV = 0.31, r^2 = 0.146, SEP = 0.34, RPD = 1.08 and bias = -0.0165. The model with the SMT spectrum was obtained R^2 = 0.415, SECV = 0.29, r^2 = 0.043, SEP = 0.35, RPD = 1.03 and bias = 0.0456. The model with the SMT+SNV spectrum obtained R^2 = 0.282, SECV = 0.32, r^2 = -0.238, SEP = 0.40, RPD = 0.90 and bias = 0.0286. The all models were poor due to r^2. Figure 7.4 shows scatter plots of reference and predicted of N concentration from model which the model using

raw spectrum (Figure 7.4a), SMT spectrum (Figure 7.4b) and SMT+SNV spectrum (Figure 7.4c).

Table 7.2 The THz model prediction results for nitrogen concentrations in durian leaf.

Pre	Calibration set			Prediction set			
treatment	nLVs	R^2	SECV	r^2	SEP	RPD	Bias
Raw	3	0.366	0.31	0.146	0.33	1.08	-0.0165
SMT	7	0.415	0.29	0.043	0.35	1.03	0.0456
SMT+SNV	3	0.282	0.32	-0.238	0.40	0.90	0.0286

* nLVs, number of latent variables (PLS factors); R^2, coefficient of determination of calibration set; SECV, standard error of cross validation; r^2, coefficient of determination of prediction set; SEP, standard error of prediction; RPD, ratio of prediction to deviation; SMT, smoothing; SMT+SNV, smoothing+vector normalization.

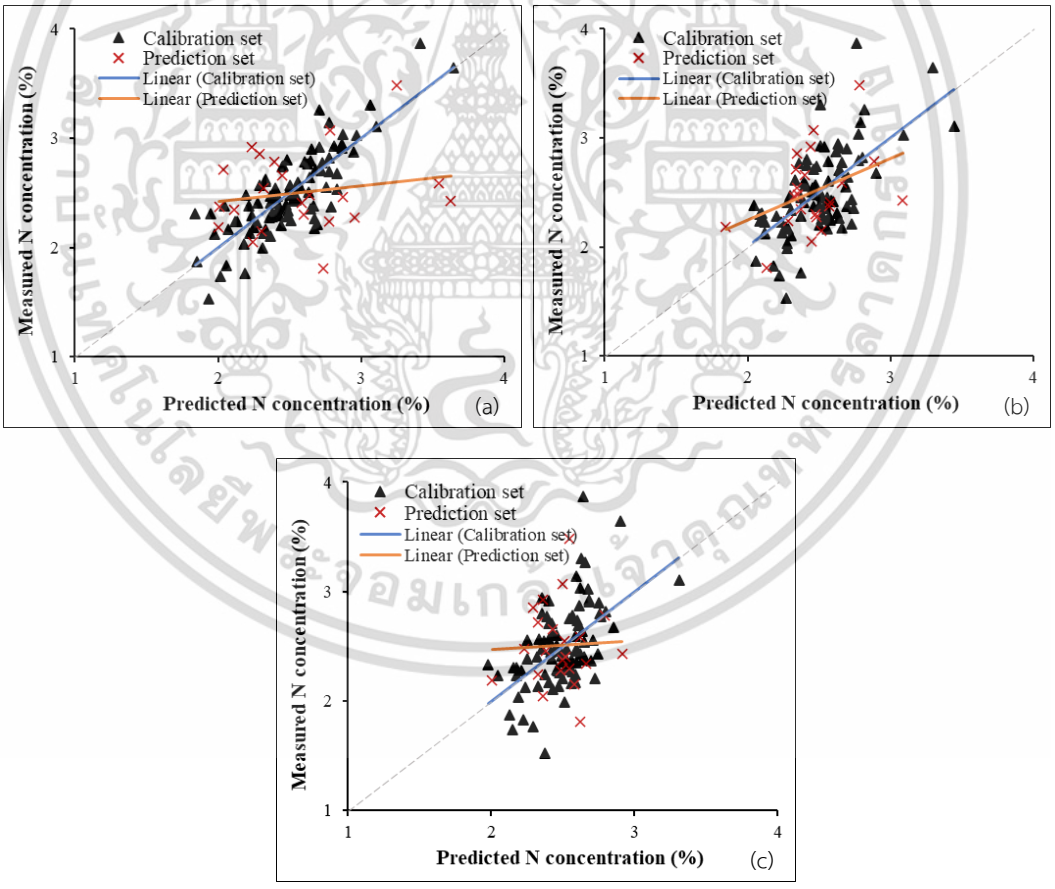


Figure 7.4 Scatter plots of measured value and predicted value to predict N content with raw (a), SMT (b), and SMT+SVN (c) spectrum.

7.4 Discussion

The pretreatment could not improve the relationship of THz absorption coefficient (**α**) and N content of durian leaf. The correlation between THz spectroscopy and N content of durian leaves at 0.71, 1.45, 1.73 and 1.86 THz was small which was not as high as in other leaves. The model's performance for N content evaluation in durian leaf by PLS models had low accuracy (r^2<0.66 and RPD<2.00). This technique was not related to the N content of durian leaves.

The comparison between NIR spectroscopy and THz spectroscopy for evaluating N content in dried ground durian leaf pellets revealed that NIR spectroscopy is more suitable for predicting N content compared to THz spectroscopy. This is because the correlation between the NIR spectrum and N concentration (the highest = 0.82) in dried ground leaf pellets is higher than that of the THz spectrum (the highest = 0.31).

7.5 Conclusion

This work describes the application of Terahertz time-domain (THz-TD) measurement to evaluate the concentrations of N in durian (*Durio zibethinus* Murray CV. Monthong) leaves. However, THz spectroscopy was unable to detect the N content of durian leaves, as no relationship was found. Consequently, the PLS models were unable to estimate the N content of durian leaves.

7.6 Reference

[1] Wang, Y., Wang, D., Shi, P., and Omasa, K. "Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light" **Plant Methods**, vol. 36. 2014. https://doi.org/10.1186/1746-4811-10-36.

[2] Zhao, D., Reddy, K. R., Kakani, V. G., and Reddy, V. R. "Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of

sorghum" **European Journal of Agronomy**, vol. 22. 2005. pp.391-403. https://doi.org/10.1016/j.eja.2004.06.005.

[3] Wagner, F. M., Melnikas, S., Cramer, J., Damry, D. A., Xia, C. Q., Peng, K., Jakob, G., Kläui, M., Kičas, S., and Johnston, M. B. "Optimised Spintronic Emitters of Terahertz Radiation for Time-Domain Spectroscopy" **Journal of Infrared, Millimeter, and Terahertz Waves**, vol. 44. 2023. pp.52-65. https://doi.org/10.1007/s10762-022-00897-9.

[4] Sousa, C. and Lopes, J. A., "Introduction and New Trends," *vol. 80*, **Comprehensive Analytical Chemistry**, 2018, pp. 1-13.

[5] Reyes-Reyes, E. S., Carriles-Jaimes, R., and Castro-Camus, E. "Algorithm for Determination of Cutoff Frequency of Noise Floor Level for Terahertz Time-Domain Signals" **Journal of Infrared, Millimeter, and Terahertz Waves**, vol. 43. 2022. pp.847-856. https://doi.org/10.1007/s10762-022-00886-y.

[6] Afsah-Hejri, L., Hajeb, P., Ara, P., and Ehsani, R. J. "A Comprehensive Review on Food Applications of Terahertz Spectroscopy and Imaging" **Compr Rev Food Sci Food Saf**, vol. 18. 2019. pp.1563-1621. https://doi.org/10.1111/1541-4337.12490.

[7] Latha, A. M., Unnikrishnakurup, S., Jain, A., Pathra, M. K., and Balasubramaniam, K. "Material Characterization and Thickness Measurement of Iron Particle Reinforced Polyurethane Multi-layer Coating for Aircraft Stealth Applications Using THz-Time Domain Spectroscopy" **Journal of Infrared, Millimeter, and Terahertz Waves**, vol. 43. 2022. pp.582-597. https://doi.org/10.1007/s10762-022-00874-2.

[8] Zahid, A., Abbas, H. T., Heidari, H., Imran, M., Alomainy, A., and Abbasi, Q. H. "Electromagnetic Properties of Plant Leaves at Terahertz Frequencies for Health Status Monitoring," presented at the 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), Nanjing, China, 2019.

[9] Shchepetilnikov, A. V., Zarezin, A. M., Muravev, V. M., Gusikhin, P. A., and Kukushkin, I. V. "Quantitative analysis of water content and distribution in plants using terahertz

imaging" **Optical Engineering**, vol. 59. 2020. https://doi.org/10.1117/1.OE.59.6.061617.

[10] Hadjiloucas, S., Karatzas, L. S., and Bowen, J. W. "Measurements of leaf water content using terahertz radiation" **IEEE Transactions on Microwave Theory and Techniques**, vol. 47. 1999. pp.142-149. https://doi.org/10.1109/22.744288.

[11] Coquillat, D., O'Connor, E., Brouillet, E. V., Meriguet, Y., Bray, C., Nelson, D. J., Faulds, K., Torres, J., and Dyakonova, N. "Terahertz Vibrational Modes of Sodium Magnesium Chlorophyllin and Chlorophyll in Plant Leaves" **Journal of Infrared, Millimeter, and Terahertz Waves**, vol. 44. 2023. pp.245-264. https://doi.org/10.1007/s10762-023-00905-6.

[12] Zhang, X., Duan, C., Wang, Y., Gao, H., Hu, L., and Wang, X. "Research on a nondestructive model for the detection of the nitrogen content of tomato" **Front Plant Sci**, vol. 13. 2022. pp.1093671. https://doi.org/10.3389/fpls.2022.1093671.

[13] Department of Agriculture. "Nutrient Management and Fertilizing Durian" [Online]. Available: https://www.doa.go.th/share/attachment.php?aid=2975. 2002.

[14] Wang, H., Kataoka, H., Tsuchikawa, S., and Inagaki, T. "Terahertz time-domain spectroscopy as a novel tool for crystallographic analysis in cellulose: cellulose I to cellulose II, tracing the structural changes under chemical treatment" **Cellulose**, vol. 29. 2022. pp.3143-3151. https://doi.org/10.1007/s10570-022-04493-x.

[15] Wold, S., Sjostrom, M., and Eriksson, L. "PLS-regression: a basic tool of chemometrics" **Chemometrics and Intelligent Laboratory Systems**, vol. 58. 2001. pp.109-130.

[16] Abdi, H., Williams, L. J., and Valentin, D. "Partial least squares regression and projection on latent structure regression (PLS Regression)" [Online]. Available: https://www.utdallas.edu/~herve/abdi-wireCS-PLS2010.pdf. 2010.

[17] Williams, P. and Norris, K. **Near-Infrared Technology: In the Agricultural and Food Industries.** American Association of Cereal Chemists. 1987.

Chapter 8

Conclusion and recommendations

8.1 Conclusion

In summary, this study aimed to address 5 objectives. The following conclusion discussed the findings from these objectives.

1) Based on the findings of this thesis, the results confirm the feasibility of applying near infrared (NIR) technique using artificial neural network (ANN) algorithm to classify the nutrient levels (low, appropriate, high) for nitrogen (N), phosphorus (P), and potassium (K) in durian (CV. Monthong) leaf, including fresh leaf and dried ground leaves.
2) The imbalanced data can adversely affect the classification model's accuracy by decreasing the accuracy for minority class samples where only small number of misclassify in minority class caused high percentage in classifying error. To address this issue and achieve more balanced data, the synthetic minority oversampling technique (SMOTE) was proved to improve the model's performance.
3) The developed regression model using partial least squares (PLS) analysis is suitable for evaluating the nitrogen (N), and potassium (K) content in durian leaf. The results indicate that the model for N and K concentration was acceptable, and the predictions were considered good. Although NIR cannot directly measure N and K in leaves, it has been found that the wavelength of protein plays a crucial role in predicting the N and K content in durian leaves. K aids in the synthesis of proteins, while N is a component of proteins. However, the model for P content did not yield successful

results due to the limited range of P content and low correlation between NIR spectra and P content in durian leaf.

4) The spectral feature wavelength selection techniques can aid in determining the most significant wavelengths for modeling. As a result, though less prediction time consumed, the model's performance did not significantly differ from the model using the full or short wavelengths.

5) The terahertz technique was not accurately determining N content in durian dried ground leaves pellets, as there was no significant correlation between the terahertz spectrum and N content. Therefore, for estimating N in durian dried ground leaves, NIR measurement is still more suitable.

This research was successful in the rapid measurement of N, P, and K concentration levels (insufficient, appropriate, and oversufficient) in durian (CV. Monthong) leaf using an FT-NIR spectrometer. The model can accurately classify the nutrient levels in fresh durian leaves. This enables farmers to save time when screening the nutrient levels in durian leaves and allows them to promptly adjust and optimize fertilizer usage, reducing costs in durian cultivation. If the result of N, P, and K estimation by NIR is insufficient, farmers may need to supply 25-30% more amount of normal fertilizer. If the result is appropriate, fertilizer can be applied as normal practice. If the result is oversufficient, the fertilizer does not need to be applied.

8.2 Recommendations

Since the R^2_{max} for P and K analysis using ICP-OES analyzer were not high, which could affect the model's performance. To increase the R^2_{max}, it is advised that the nutritional analysis methods should be changed. For instance, P content can be analyzed using the Macro method (AOAC 933.01-1933), and K can be analyzed using the Gravimetric

method (AOAC 925.01-1925), etc. These alternative methods help to improve the accuracy and reliability of the model performance.

The improvement of the K content level evaluation model should include more samples with high K concentration because currently there is a lack of data on high K content for the classification model that covers the full concentration range of K. Therefore, treatment involving foliar fertilizer to prepare samples with high concentrations is necessary. This will allow for an improvement in the model's performance.

To improve the performance of the P content evaluation model, the range of sample content should be expanded widely by foliar fertilizer. The available P data is currently very limited and covers a narrow range, which affects the model's performance negatively.

Additionally, exploring the creation of a non-linear modeling could enhance the model's performance. Therefor the non-line modeling should be applied to quantitatively determination of N, P and K content in durian leaves. Similarly, increasing the number of N and K samples would also make the model more robust by covering a wider range and variation of data.

Lastly, based on the findings of this research, which indicated the protein mostly affect to N and K content prediction. The protein content in durian leaves is recommended to be measured for assessment the correlation between N or K and protein in order to support that NIR models can predict N and K content in leaf via protein content.

Appendix

Published papers

International Published papers:

1. Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling" **Measurement.** vol.203. 2022. pp.111975.
2. Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Classification of N, P, and K concentrations in durian (*Durio Zibethinus* Murray CV. Mon Thong) leaves using near-infrared spectroscopy" **Engineering and Applied Science Research.** vol.49(1), 2022. pp.127-132.

Oral presentation:

1. Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Overall Precision Test for Determination the Nutrient in Durian Leaf in Durian Orchard using Near-Infrared Spectroscopy" **The 7th Asian NIR Symposium conference (ANS2020).** at Avani Hotel in Khonkaen, Thailand. February 12-15, 2020.
2. Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Classification of N, P, and K concentrations in durian (*Durio Zibethinus* Murray CV. Mon Thong) leaves using near-infrared spectroscopy" **14th Thai Society of Agricultural Engineering International Conference (TSAE 2021)**, online, Thailand. May 12-13, 2021. (Presentation award).
3. Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "The Near Infrared Spectroscopy Model Lowest Limit of Detection and Quantification for Nutrient in Durian (CV Monthong) Leaf Evaluation" **The 8th Asian NIR Symposium conference (ANS2022).** online, South korea. November 28-30, 2022.

Author biography

Name-Surname Miss Thitima Phanomsophon

Date of birth 14 September 1994

Contact address 16/13, moo.8, Nongplalai, Banglamung, Chonburi, 20150

Email address Thitimap.june@gmail.com

Contact number 091-8628484

Education background

D.Eng. (Agricultural Engineering) from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand (2019-2023, GPA: 4.00)

M.Eng. (Agricultural Engineering) from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand (2017-2018, GPA: 3.88)

B.Eng. (Agricultural Engineering, 2nd class honor) from King Mongkut's Institute of Technology Ladkrabang, Thailand (2013-2017, GPA: 3.43)

Training attended

- Visiting researcher in Graduate School of Bio agricultural Sciences, Nagoya University, Japan (6 months, 2022)
- Summer training program at Port Authority of Thailand, Bangkok, Thailand (2 month, 2016)

Scholarships

- Royal Golden Jubilee scholarship (RGJ) PhD program from National Research Council of Thailand (NRCT) [Grant numbers NRCT5-RGJ63021-169], 2020.

- The Agricultural Research Development Agency (Public Organization) [grant number PRP6305031290], 2020.

Oral presentation

- Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "The Near Infrared Spectroscopy Model Lowest Limit of Detection and Quantification for Nutrient in Durian (CV Monthong) Leaf Evaluation" **The 8th Asian NIR Symposium conference (ANS2022).** online, Korea. November 28-30, 2022.
- Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Classification of N, P, and K concentrations in durian (*Durio Zibethinus* Murray CV. Mon Thong) leaves using near-infrared spectroscopy" **14th Thai Society of Agricultural Engineering International Conference (TSAE 2021).** online, Thailand. May 12-13, 2021. (Presentation award).
- Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Overall Precision Test for Determination the Nutrient in Durian Leaf in Durian Orchard using Near-Infrared Spectroscopy" **The 7th Asian NIR Symposium conference (ANS2020).** at Avani Hotel in Khonkaen, Thailand. February 12-15, 2020.
- Phanomsophon, T., and Sirisomboon, P. "Precision test for scanning of soil in durian orchard using near-infrared spectroscopy" **12th Thai Society of Agricultural Engineering International Conference (TSAE 2019).** At Hard rock Pattaya, Chonburi, Thailand. March 14-15, 2019.
- Phanomsophon, T., and Sirisomboon, P. "Precision Test for Monitoring of Acetification Process in Vinegar Fermented Bioreactor using Near-Infrared Spectroscopy" **10th Thai Society of Agricultural Engineering International**

Conference (TSAE 2017). At Impact exhibition center, Bangkok, Thailand. September 7-9, 2017.

International Published papers

- Pornchaloempong, P., Sharma, S., Phanomsophon, T., Srisawat, K., Inta, W., Sirisomboon, P., Prinyawiwatkul, W., Nakawajana, N., Lapcharoensuk, R., and Teerachaichayut, S. "Non-Destructive Quality Evaluation of Tropical Fruit (Mango and Mangosteen) Purée Using Near-Infrared Spectroscopy Combined with Partial Least Squares Regression" **Agriculture.** vol.12(12). 2022. pp.2060.
- Phanomsophon, T., Jaisue, N., Worphet, A., Tawinteung, N., Shrestha, B., Posom, J., Khurnpoon, L., and Sirisomboon, P. "Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling" **Measurement.** vol.203. 2022. pp.111975.
- Phanomsophon, T., Jaisue, N., Tawinteung, N., Khurnpoon, L., and Sirisomboon, P. "Classification of N, P, and K concentrations in durian (*Durio Zibethinus* Murray CV. Mon Thong) leaves using near-infrared spectroscopy" **Engineering and Applied Science Research.** vol.49(1), 2022. pp.127-132.
- Saechua, W., Sharma, S., Nakawajana, N., Leepaitoon, K., Chunsri, R., Posom, J., Roeksukrungrueang, C., Siritechavong, T., Phanomsophon, T., Sirisomboon, P., Lapcharoensuk, R., and Pornchaloempong, P. "Integrating Vis-SWNIR spectrometer in a conveyor system for in-line measurement of dry matter content and soluble solids content of durian pulp" **Postharvest Biology and Technology.** vol.181. 2021. pp.111640.
- Phanomsophon, T., and Sirisomboon, P. "Precision test for scanning of soil in durian orchard using near-infrared spectroscopy" **IOP Conference Series: Earth and Environmental Science.** vol.301. 2019. pp.012070.

- Phanomsophon, T., Sirisomboon, P., Lapcharoensuk, R., Shrestha, B., and Krusong, W. “Evaluation of acetic acid and ethanol concentration in a rice vinegar internal venturi injector bioreactor using Fourier transform near infrared spectroscopy” **Journal of Near Infrared Spectroscop**y. vol.27(6). 2019. pp.416-423.

Areas of expertise and interest

- Application of near infrared spectroscopy
- Utilization of non-destructive testing techniques for agricultural products and food
- Model analysis
- Writing code by MATLAB program

Reference persons

- Prof. Dr. Panmanas Sirisomboon, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Thailand
  Email: panmanas.si@kmitl.ac.th