ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Ripeness Evaluation Using Near-Infrared (NIR) Spectroscopy and NLP for Interpretability

Panudech Tipauksorn[1], Jutturit Thongpron[2], Prasert Luekhong[3], Minoru Okada[4] and Krisda Yingkayun[5]

ABSTRACT

This study develops a non-destructive avocado ripeness classification system using a low-cost, portable near-infrared (NIR) scanner and machine learning. Traditional ripeness assessment methods are often destructive and subjective, limiting their efficiency in agricultural practices. To address this issue, we developed a custom NIR scanner capable of capturing spectral information across 18 discrete wavelength bands for avocado ripeness classification. The research focuses on Buccaneer avocados sourced from the Royal Project, with samples collected from both the Royal Project Gardens and Sorting Plant. A total of 120 kg of avocados were systematically sampled and categorized by agricultural experts into three ripeness stages :raw, ripe, and aged. This study applies Multiplicative Scatter Correction (MSC) to preprocess NIR spectra, enhancing feature separation before machine learning model training. This study evaluates three classification models: Random Forest, XGBoost, and the Gaussian Mixture Model (GMM). Random Forest achieved the highest classification accuracy (78%) with an AUC score of 0.93, followed by XGBoost (74% accuracy, AUC 0.91). GMM performed worse, with 42% accuracy and an AUC of 0.58. Additionally, Natural Language Processing (NLP) was applied to convert model predictions into human-readable ripeness descriptions, assisting farmers in decision-making. This study demonstrates that low-cost NIR technology combined with AI-driven analysis enables efficient, non-destructive classification of avocado ripeness.

Article information:

Keywords: Near-Infrared Spectroscopy (NIR), Machine Learning for Ripeness Classification, Natural Language Processing (NLP) for Model Interpretability

Article history:
Received: February 30, 2025
Revised: October 9, 2025
Accepted: December 4, 2025
Published: January 10, 2026 (Online)

DOI: 10.37936/ecti-cit.2026201.260912

1. INTRODUCTION

Accurately determining fruit ripeness is essential for optimizing harvest timing, reducing postharvest losses, and ensuring product quality. Traditional ripeness evaluation methods rely on destructive testing and subjective assessments, which are inefficient for large-scale applications. There is a growing need for non-destructive, objective, and automated approaches to improve efficiency in agricultural production.

Researchers commonly use NIR spectroscopy for non-destructive classification of fruit ripeness. While they work well, many models are still hard to understand. Tipauksorn et al. [26] used comparable low-cost sensors to perform simple binary classification with color LEDs that required user interpretation. This study extends previous research by examining three ripeness stages and integrating NLP to provide more precise feedback. This integration improves interpretability and usability, bridging the gap between advanced NIR systems and practical tools for farmers and stakeholders in agricultural decision-making.

This research aims to develop a machine learning-based avocado ripeness classification system using NIR spectroscopy and Natural Language Processing (NLP) for interpretability.

[1,2,5]The authors are with the Department of Electrical Engineering, Faculty of Engineering, Rajamangala University of Technology Lanna, Thailand, Email: panudech@rmutl.ac.th, jutturit@rmutl.ac.th and krisda@rmutl.ac.th
[3]The author is with the College of Integrated Science and Technology, Rajamangala University of Technology Lanna, Thailand, Email: prasert@rmutl.ac.th
[4]The author is with the Division of Information Science, Nara Institute of Science and Technology, Japan, Email: mokada@is.naist.jp
[5]Corresponding author: krisda@rmutl.ac.th

This study sampled 120 kg of avocados and collected NIR spectra across 18 wavelength bands. The dataset was processed using Multiplicative Scatter Correction (MSC) to reduce spectral noise, and models including Random Forest, XGBoost, and Gaussian Mixture Model (GMM) were trained and evaluated for classification performance. Additionally, this study implements NLP to translate model predictions into human-readable ripeness descriptions.

This study contributes to AI-driven agricultural monitoring by combining NIR spectroscopy, machine learning, and NLP to improve ripeness evaluation accuracy and interpretability. The results support improved decision-making for farmers, postharvest management, and agricultural supply chain optimization.

2. LITERATURE REVIEW

Accurate and efficient ripeness evaluation is critical for optimizing harvest timing, minimizing postharvest losses, and ensuring high-quality fruit products. Traditional methods, often relying on destructive testing or subjective sensory evaluations, are time-consuming, labour-intensive, and lack objectivity [1], [2]. Near-infrared (NIR) spectroscopy emerges as a powerful non-destructive alternative, offering rapid and objective assessments based on the unique spectral fingerprint of each fruit [3], [4]. However, interpreting the complex NIR spectral data and developing robust predictive models demands sophisticated chemometric techniques and increasingly, the application of artificial intelligence (AI), including deep learning approaches [5], [6]. This literature review examines the application of NIR spectroscopy in fruit ripeness evaluation, focusing on the role of natural language processing (NLP) in enhancing model interpretability and understanding.

NIR spectroscopy exploits the interaction between near-infrared light (typically 780–2500 nm) and molecular vibrations within a substance [7], [8]. Different fruit components, including sugars, acids, and water, exhibit unique absorption or reflectance patterns at specific wavelengths. The resulting spectral data, a collection of absorbance or reflectance values across the NIR spectrum, serves as a fingerprint reflecting the fruit chemical composition and ripeness stage [3], [9]. This non-destructive technique allows for rapid, in-situ analysis, minimizing sample preparation and avoiding the need for destructive sampling methods [10], [11].

The applications of NIR spectroscopy in fruit ripeness assessment are diverse and expanding. Studies have successfully employed this technique to predict total soluble solids (TSS) in strawberries [9], assess the quality of mangoes [6], and evaluate the internal quality of various citrus fruits, including oranges, lemons, clementines, tangerines, and Tahiti limes [3]. Furthermore, researchers have explored its use in watermelon ripeness detection [4], demonstrating its potential for broader application across various fruit types. Researchers have applied this versatile technology beyond fruit analysis, including wheat quality assessment [12], nicotine content evaluation in tobacco [13], and post-consumer textile waste classification [14].

Raw NIR spectra are often highly complex and not directly interpretable. Chemometric methods play a crucial role in extracting meaningful information and building predictive models that relate spectral features to ripeness [5], [6]. Traditional chemometric techniques, such as multiple linear regression (MLR) and partial least squares regression (PLSR), have been widely utilized [3], [9]. However, these methods can often lack transparency, making it difficult to understand the specific relationships between spectral features and ripeness [6].

Machine learning (ML) algorithms, particularly artificial neural networks (ANNs) and convolutional neural networks (CNNs), have demonstrated superior performance in NIR spectral analysis compared to traditional methods [6], [8]. These algorithms capture complex, non-linear relationships between spectral data and ripeness, which leads to improved prediction accuracy [15], [13]. For instance, CNNs have shown promise in estimating the ripening state of Fuji apples [15] and predicting sugar and pH levels in wine grapes [16]. The application of Support Vector Machines (SVM) in a watermelon ripeness detector achieved 92.5% accuracy [4], highlighting the effectiveness of ML in this domain. A lightweight 1D-CNN model demonstrated success in predicting nicotine levels in tobacco using NIR spectroscopy [13], indicating the adaptability of these techniques across different agricultural products.

Deep learning (DL), a subset of ML, offers potent tools for NIR spectral analysis [6]. DL models, with their multiple layers and ability to learn intricate features, often surpass traditional ML methods in predictive accuracy [15], [16]. The use of DL models, such as ResNet, has been successful in automatically extracting and identifying features from Fourier Transform Near-Infrared (FT-NIR) spectroscopy data for the geographical traceability of medicinal plants [17]. Researchers have applied one-dimensional convolutional neural networks (1D-CNNs) and three-dimensional ResNet architectures to determine maturity and soluble solids content in strawberries using hyperspectral imaging [18]. However, the increased complexity of DL models presents a significant challenge to interpretability, making it difficult to understand the decision-making processes underlying their predictions.

The black box nature of many ML and DL models is a significant obstacle to their wider adoption. Understanding the factors that drive model predictions remains crucial, particularly in high-stakes applications such as food safety and quality control. NLP techniques [19] offer potential solutions to improve model interpretability by bridging the gap between complex model outputs and human understanding [6].

Several NLP techniques enhance the interpretability of NIR spectral models. Attention mechanisms, commonly used in NLP for tasks like machine translation [20] and text summarization, can be adapted to highlight the most relevant wavelengths or spectral regions contributing to a ripeness prediction [6]. These techniques visually represent the model decision-making process, enabling researchers to identify the key spectral features most influential in predicting ripeness.

Natural language generation (NLG) offers another promising avenue. Natural Language Generation (NLG) creates human-readable reports that summarize model predictions, including the predicted ripeness level, confidence score, and relevant spectral features [6]. These reports facilitate communication among model developers, stakeholders, and end users, thereby promoting trust and informed decision-making. While the current literature does not extensively detail specific NLP applications in this context, the potential for such applications is clearly significant.

Effective data pre-processing is essential for successful NIR spectral analysis. Various techniques, including standard normal variate (SNV), variable sorting normalization (VSN), and Savitzky-Golay smoothing, are employed to reduce noise, correct for scattering effects, and enhance relevant spectral features [9], [21]. However, it is crucial to recognize that some pre-processing methods can negatively impact model performance, especially when both chemical and physical properties are essential for understanding fruit ripeness [21]. Careful consideration of the pre-processing method is necessary to optimize both model accuracy and interpretability.

Feature selection techniques enhance interpretability by reducing data dimensionality while preserving or improving predictive performance. Methods like competitive adaptive reweighted sampling (CARS) and variable importance in projection (VIP) help identify the most relevant wavelengths for predicting specific fruit quality attributes [17]. By focusing on a smaller subset of key spectral features, it becomes easier to understand the model predictions and the relationships between spectral data and fruit ripeness.

Despite its considerable potential, NIR spectroscopy still faces several challenges. Building robust and accurate models requires large, high-quality datasets representing the variability in fruit characteristics across different varieties, growing conditions, and harvesting seasons [6], [16]. Data scarcity can limit the generalizability of models and hinder their practical application. The high dimensionality of NIR spectra necessitates effective feature selection and dimensionality reduction techniques to prevent overfitting and improve model interpretability [6], [8].

Recent research on stereo vision-based turn-alignment optimization has introduced novel approaches to enhancing real-time spatial positioning for wireless power transmission [22]. This method enables precise spatial calibration, which benefits agricultural sensing applications such as NIR-based ripeness evaluation. By integrating stereo vision techniques with NIR spectroscopy, future research can explore enhanced accuracy in fruit quality assessment by improving sensor alignment and fruit positioning. These developments contribute to the broader goal of advancing automated agricultural monitoring and improving innovative farming applications.

Near-Infrared (NIR) spectroscopy, combined with Natural Language Processing (NLP), enables non-destructive ripeness evaluation by converting spectral data into interpretable insights. Advances in AI and IoT further optimize these methods, facilitating real-time monitoring in smart agriculture. The integration of AI-driven algorithms with high-capacity IoT networks, as discussed in [23], underscores the potential of RoF (Radio-over-Fiber) systems for efficient data transmission.

Future research should prioritize the development of more sophisticated NLP techniques for interpreting NIR spectral models, leading to a deeper understanding of their decision-making processes. Integrating NLP with explainable AI (XAI) methods could further enhance model transparency and build trust in automated ripeness assessments [6]. Exploring hybrid models that combine NIR spectroscopy with other sensing technologies, such as hyperspectral imaging and electronic noses, holds significant promise for improving the accuracy and comprehensiveness of ripeness evaluations [24], [25]. Developing portable and affordable NIR spectrometers plays a critical role in enabling broader adoption of this technology across diverse agricultural settings. The successful integration of these advanced technologies with existing agrarian practices is vital for maximizing their impact.

Integrating Near-Infrared (NIR) spectroscopy with modern machine learning and Natural Language Processing (NLP) offers a game-changing approach for assessing fruit ripeness without causing damage. Deep learning models such as CNNs achieve high classification accuracy; however, limited interpretability can hinder adoption by end users, including farmers. This study aims to improve transparency by using NLP-based explanations designed for real-world applications. Earlier work by Tipauksorn *et al.* [26] applied the same low-cost 18-band NIR hardware to binary avocado classification (raw vs. ripe) using LED outputs, whereas this study extends the task to three ripeness classes (raw, harvestable, and ready-to-eat) and integrates NLP-based interpretability to improve usability. This work addresses a critical gap in practical human–AI collaboration for agricultural decision-making.

Recent studies from 2022 to 2025 highlight the growing significance of NIR–AI integration in precision agriculture for fruits like mango, kiwi, and avocado. Many studies focus on predictive accuracy but overlook model explainability and practical usability. We address this issue by emphasizing user-friendly interpretability and immediate applicability, particularly in low-resource environments. Future research should emphasize multi-seasonal datasets, sensor fusion strategies, and multilingual NLP interfaces to encourage broader adoption across diverse farming contexts. The combination of explainable AI and portable NIR sensors has excellent potential to enhance postharvest quality monitoring, minimize food loss, and promote data-driven farming in line with global agricultural modernization trends.

3. METHODOLOGY

This study utilized near-Infrared (NIR) spectroscopy and machine learning models to classify avocado ripeness into raw, harvestable, and ready-to-eat categories. This study samples 120 kg of avocados and acquires NIR spectra across 18 wavelength bands. MSC pre-processing reduces spectral noise and improves feature separation before training Random Forest, XGBoost, and Gaussian Mixture Model (GMM) classifiers. The system further integrates NLP to translate model predictions into human-readable ripeness descriptions for agricultural decision-making.

3.1 System diagram

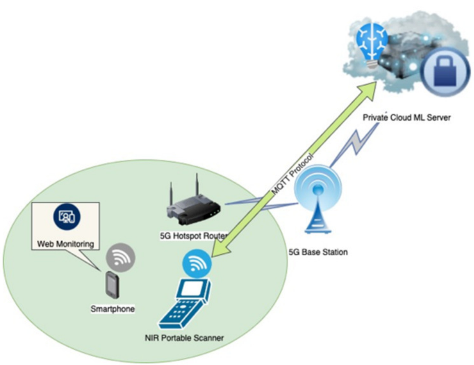

Fig. 1: *System Architecture for Cloud-Based NIR Spectroscopy Ripeness Prediction Using 5G and MQTT Protocol.*

The proposed system integrates Near-Infrared (NIR) spectroscopy with cloud-based machine learning (ML) to classify avocado ripeness in real time. It enables wireless data transmission, remote computation, and result visualization, delivering an efficient and scalable solution for ripeness assessment in agricultural environments, as shown in Figure 1.

The portable NIR scanner acquires spectral data from avocados and transmits the values to a 5G hotspot router using MQTT over a Wi-Fi connection. The router acts as a gateway, forwarding the data to a 5G network infrastructure for seamless transmission to a Private Cloud ML Server. Once the data reaches the cloud server, the machine learning model processes the NIR spectral data, classifies the ripeness stage, and generates a prediction result.

After processing, the system sends the predicted ripeness results back to the portable NIR scanner and the smartphone web monitoring application via the MQTT protocol. The raw spectral data and computed results are securely stored in the cloud server, ensuring proper data management for future analysis. Users can access real-time ripeness status and harvesting recommendations through a web-based monitoring system, improving decision-making for postharvest management.

The proposed architecture integrates 5G, cloud computing, and MQTT communication to deliver a scalable, efficient, and accessible real-time solution for intelligent agricultural monitoring.

3.2 Population and Sampling

This study examines avocados grown in a 1,600-m^2 (1 Rai) garden. To ensure accurate representation, the research assumes that 1 kilogram of avocados reflects the average yield of an avocado tree within this area. The sampling process was structured to include avocados at different ripeness stages, as outlined in Table 1.

Table 1: *Population and Sample Distribution.*

Ripeness Category	Sampled Weight (kg)	Estimated No. of Avocados
Raw	40	120–160
Ripe (Stored 7 Days)	40	120–160
Aged (Stored 3 Days)	40	120–160
Total	120	360–480

This study collects avocado samples from the Royal Project and categorizes them into three ripeness groups. Agricultural academics carefully sorted the avocados into the following categories

- Raw – Unripe avocados, not yet suitable for consumption.
- Ripe (Stored for 7 Days) – Avocados that are harvested and stored for a week before being ready-to-eat.
- Aged (Stored for 3 Days) – Avocados that ripen rapidly and require consumption within three days.

Typically, three to four avocados weigh approximately 1 kilogram, which served as a reference in the sampling process. A total of 120 kilograms of avocados were collected, ensuring a diverse and statistically significant dataset for analysis.

Figure 2 illustrates the avocado sampling process conducted as part of this study. The avocados, identified as the Buccaneer variety, were sourced from the Royal Project, specifically from its gardens and sorting plant. The top-left image presents freshly harvested avocados labeled for sample tracking.

Fig.2: *Sampling Process of Buccaneer Avocados in the Royal Project.*

The top-right image illustrates the sorting and categorization process at the Royal Project Sorting Plant, where agricultural staff classify avocados by ripeness stage. The bottom row shows key fieldwork activities using a low-cost portable NIR scanner that transmits spectral data over Wi-Fi through a 4G/5G hotspot, including harvesting, initial quality inspection, and transport to the sorting facility. This systematic sampling ensures representative data collection for the study.

3.3 Data Collection and Variable Measurement

The study utilizes near-Infrared (NIR) spectroscopy to analyze the spectral properties of avocados at different ripeness stages. The system stores the collected data in a database for subsequent processing and machine learning–based classification.

Data Collection Process

- Avocado Selection – Avocados are categorized based on ripeness.
- Spectral Scanning – Each sample undergoes NIR spectroscopy, capturing data at 18 wavelengths.
- Data Storage – The spectral information is stored in a structured dataset.
- Pre-processing – The raw spectra undergo Multiplicative Scatter Correction (MSC) to reduce noise.

3.4 Data Analysis and Statistical Methods

The research involves machine learning-based classification to distinguish between different ripeness categories. The analysis follows these steps.

Table 2: *Variable Definitions.*

Variable Name	Description	Measurement Unit
Spectrum1–Spectrum18	NIR spectral features measured at 18 wavelengths	Absorbance/ Reflectance
Evaluate the Type	Ripeness category of the avocado	Categorical (Raw, Ripe, Aged)

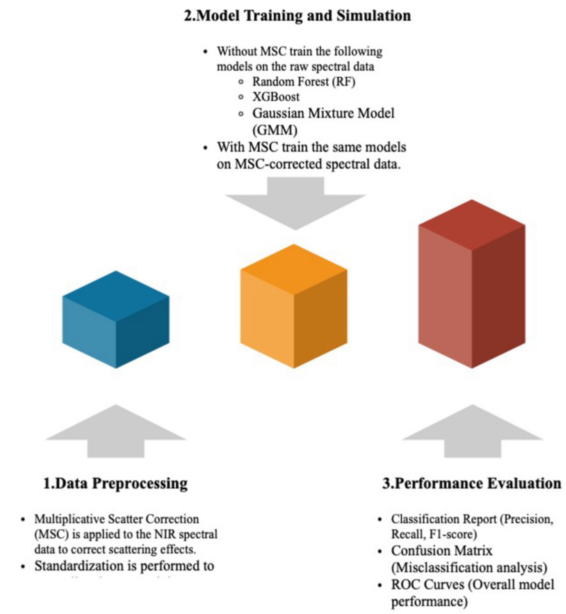

Fig.3: *Shows data analysis and statistical process step.*

This Figure 3 illustrates the Data Analysis and Statistical Methods workflow in three main steps

1. Data Pre-processing
   - Multiplicative Scatter Correction (MSC) is applied to the Near-Infrared (NIR) spectral data to correct scattering effects and improve signal quality.
   - The preprocessing stage applies standardization to normalize spectral data and ensure consistent input values for machine learning models.
2. Model Training and Simulation
   - Without the MSC training machine learning models on the raw spectral data, including
     - Random Forest (RF)
     - XGBoost
     - Gaussian Mixture Model (GMM)
   - With MSC, train the same models using MSC-corrected spectral data to observe performance improvements.
3. Performance Evaluation
   - The study assesses the trained models using three key evaluation metrics.

- Classification Report Measures Precision, Recall, and F1-score for each ripeness category.
- A confusion matrix analyzes the number of correct and incorrect classifications.
- ROC Curves evaluate the model ability to distinguish between different ripeness categories.

3.5 Equation for Classification Accuracy

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \tag{1}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

3.6 Ripeness Interpretation to Natural Language

After the system determines ripeness, it converts the classification results into human-readable text to support decisions on fruit harvest readiness. The system derives the interpretation from a ripeness score calculated using a weighted confidence approach based on the classification model output probabilities.

1. Ripeness Score Calculation

The system computes the ripeness score using the formula in (2).

$$Ripeness_{Score} = \frac{CRipe}{CRaw + CRipe + CAged} \times 100 \tag{2}$$

Where:

- $CRaw$ = The model probability output for the fruit being raw.
- $CRipe$ = The model probability output for the fruit being ripe for a week before eating.
- $CAged$ = The model probability output for the fruit ready-to-eat in 3 days.

This score quantifies the ripeness level on a scale from 0 to 100, ensuring an objective interpretation of the spectral data.

2. Table 3 summarizes the interpretation based on the ripeness score.

By applying this approach, the study enhances the practical usability of the classification model, allowing farmers, agricultural experts, and supply chain managers to make informed decisions regarding optimal harvesting time.

This visualization helps explain how the system derives the ripeness score from model confidence values, ensuring an objective and data-driven interpretation of ripeness.

Table 3: *To provide meaningful insights, the system categorizes the calculated ripeness score into three levels.*

Ripeness Score (%)	Interpretation
0 – 40%	raw and not ready for harvest.
41 – 80%	harvestable and can be stored for a week before eating.
81 – 100%	ready-to-eat and should be consumed within 3 days.

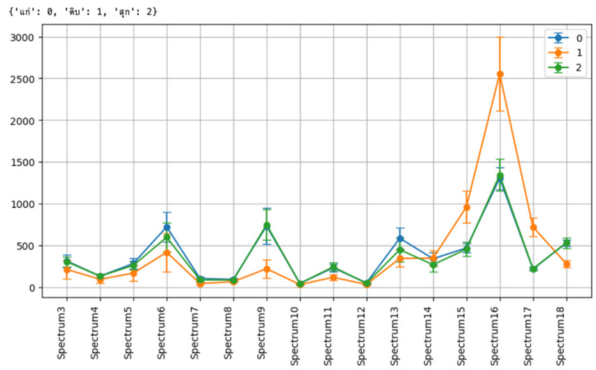

Fig.4: *Derivation of the ripeness score from model confidence values.*

The graph in Figure 4 visually represents the relationship between ripeness score, model confidence levels, and interpretation categories. The black curve shows the calculated ripeness score, derived from the confidence values of different ripeness stages. The dashed lines indicate the probability trends for Raw (blue), Ripe (orange), and Aged (red) classifications. As the fruit ripens, raw confidence decreases, ripe confidence increases, and aged confidence gradually rises. The graph separates ripeness into three shaded regions: 0–40% indicates raw fruit, 41–80% indicates a maturing stage suitable for near-term harvest, and 81–100% confirms fully ripe fruit ready for harvest. This structured visualization enables a clear, data-driven interpretation of fruit ripeness, facilitating informed decision-making for harvest timing.

4. RESULTS AND DISCUSSION

Figure 5 presents the Near-Infrared (NIR) spectral responses of avocados grouped into three ripeness stages: raw, ripe, and aged. The x-axis shows the 18 wavelength bands measured by the NIR sensor, while the y-axis indicates the corresponding absorbance intensity. The data reveal distinct spectral variations among ripeness categories, with pronounced differences at specific wavelengths.

A key observation is that ripe avocados (category 2, green line) generally exhibit higher spectral intensity in the mid-wavelength region, indicating increased molecular changes associated with ripening. Aged avocados (category 0, blue line) show similar patterns but with slight deviations in peak intensities, suggesting progressive biochemical transformations. Raw avocados (category 1, orange line) show significantly higher absorbance at specific wavelength bands (e.g., Spectrum16), likely due to unripe tissue composition and higher moisture content.

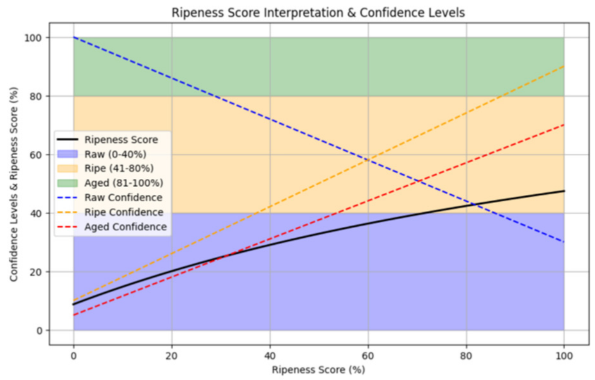

Fig.5: *Spectrum of NIR from Avocado Dataset Across Three Ripeness Categories.*

The spectral trends confirm that specific NIR wavelengths are highly sensitive to biochemical differences in ripeness, reinforcing the suitability of NIR spectroscopy for non-destructive ripeness classification. These findings support further investigation into optimal feature selection for machine learning-based ripeness prediction models.

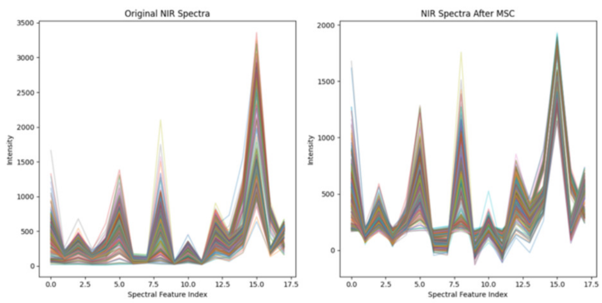

Fig.6: *Comparison of Original and MSC-Processed NIR Spectra Before Model Training.*

Figure 6 presents a comparative analysis of the Near-Infrared (NIR) spectral data before and after applying Multiplicative Scatter Correction (MSC), a crucial pre-processing step in spectral analysis. The left graph displays the raw spectral data, while the right graph illustrates the corrected spectra after MSC pre-processing.

The original NIR spectra exhibit substantial intensity variations across samples, accompanied by pronounced spectral scattering. Factors such as light scattering, variations in sample thickness, and surface irregularities contribute to these inconsistencies and introduce unwanted noise into the dataset. Without preprocessing, such variations increase model complexity and degrade machine learning classification performance.

After applying MSC, the right graph demonstrates a significant improvement in spectral consistency. The MSC method reduces baseline variations and normalizes intensity differences, leading to more uniform spectral responses across samples. This preprocessing step improves the extraction of relevant spectral features by minimizing scattering artifacts, thereby improving the reliability of the input data for classification models. The improved spectral alignment achieved through MSC highlights its importance in enhancing model generalization and reducing noise in spectral-based ripeness classification. These findings confirm that applying MSC before model training enhances the predictive power of machine learning algorithms, making it a necessary step in non-destructive fruit quality assessment.

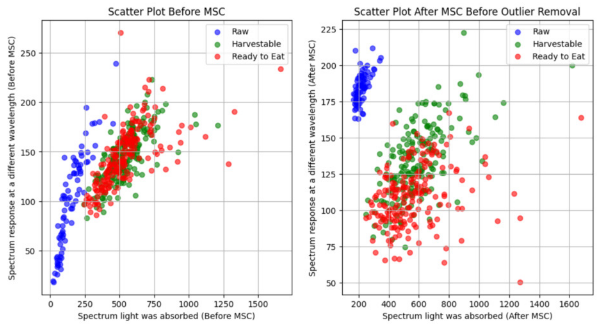

Fig.7: *Scatter Plot Comparison Before and After MSC Pre-processing.*

Figure 7 illustrates the scatter distribution of NIR spectral data across three ripeness categories (Raw, Harvestable, and Ready-to-eat) before and after applying Multiplicative Scatter Correction (MSC). The left scatter plot represents the raw spectral data without MSC pre-processing, while the right plot shows the spectral distribution after MSC pre-processing but before outlier removal.

In the scatter plot before MSC (left graph), the spectral data points exhibit significant clustering with overlapping distributions, particularly between the Harvestable (green) and Ready-to-eat (red) categories. The presence of spectral distortions due to light scattering and sample inconsistencies results in less distinguishable class separations, which can reduce the effectiveness of machine learning classifiers.

After applying MSC (right graph), the spectra show a more consistent and well-aligned distribution. The Raw category (blue) remains well-separated, while the separation between Harvestable and Ready-to-eat categories is enhanced, providing a more structured dataset for classification. In addition, the spectral data show a more uniform distribution, indicating that MSC effectively reduces baseline variation and corrects scattering effects, resulting in better-defined clusters.

The improved spectral alignment achieved through MSC pre-processing enhances the discriminability of ripeness categories, thereby improving classification accuracy in machine learning models. These findings confirm the importance of MSC preprocessing in non-destructive fruit quality assessment by minimizing spectral variation prior to model training.

Figure 8 shows the scatter plot of NIR spectral data following the application of Multiplicative Scatter Correction (MSC) and improved outlier removal. This version effectively maintains all three ripeness categories Raw (blue), Harvestable (green), and Ready-to-eat (red) demonstrating the success of adjusting the outlier detection threshold.

During preprocessing, the pipeline mistakenly removed the entire raw category because its distinct spectral characteristics and lower intensity ranges resembled outliers. Strict parameter settings in the Interquartile Range (IQR) method or anomaly detection models such as Isolation Forest caused this issue.

Relaxing the IQR threshold or using feature-informed anomaly detection allows for the retention of raw samples with unique yet valid spectral signatures. This change boosts dataset representativeness and strengthens model robustness.

The updated scatter distribution (right graph) shows clear class separability, with distinct spatial grouping for all three ripeness stages. The Raw category now stands out as a unique cluster, highlighting its distinct spectral properties and the importance of careful outlier management in spectral datasets.

Fig.8: *Scatter Plot After MSC with Outlier Removal.*

Additionally, the observed spectral spread between the Harvestable and Ready-to-eat categories remains well-structured, confirming that MSC pre-processing and outlier removal improve class separability. The complete removal of the raw category indicates that the preprocessing thresholds require adjustment or that alternative outlier detection methods should be considered to retain meaningful raw avocado data while removing true anomalies. This outcome highlights the importance of careful outlier handling in spectral analysis to preserve class-specific information and improve data quality for machine learning–based ripeness classification.

Figure 9 presents the classification reports of three machine learning models Random Forest, XGBoost, and Gaussian Mixture Model (GMM) used for avocado ripeness classification. This study evaluates the models using precision, recall, F1-score, and accuracy to assess their effectiveness in distinguishing harvestable, raw, and ready-to-eat categories.

This study selects the three models for their ability to handle classification tasks involving spectral data.

1. Random Forest (RF) – A robust ensemble learning algorithm that effectively handles non-linear patterns and is widely used in agricultural classification problems.
2. XGBoost (Extreme Gradient Boosting) – An optimized gradient boosting algorithm that enhances predictive performance with improved handling of imbalanced data.
3. Gaussian Mixture Model (GMM) – A probabilistic clustering valuable method for unsupervised learning and identifying distinct spectral patterns.

```
📊 Classification Report Random forest model:
              precision    recall  f1-score   support

 Harvestable       0.81      0.74      0.78        47
         Raw       1.00      1.00      1.00         5
Ready to Eat       0.72      0.79      0.76        39

    accuracy                           0.78        91
   macro avg       0.84      0.85      0.84        91
weighted avg       0.78      0.78      0.78        91

📊 Classification Report XGBoost model:
              precision    recall  f1-score   support

 Harvestable       0.77      0.72      0.75        47
         Raw       0.83      1.00      0.91         5
Ready to Eat       0.68      0.72      0.70        39

    accuracy                           0.74        91
   macro avg       0.76      0.81      0.79        91
weighted avg       0.74      0.74      0.74        91

📊 Classification Report Gaussian Mixture model:
              precision    recall  f1-score   support

 Harvestable       0.47      0.57      0.51        47
         Raw       1.00      1.00      1.00         5
Ready to Eat       0.29      0.21      0.24        39

    accuracy                           0.44        91
   macro avg       0.58      0.59      0.58        91
weighted avg       0.42      0.44      0.42        91
```

Fig.9: *Classification Performance of Random Forest, XGBoost, and Gaussian Mixture Models.*

Comparison of Classification Results

- Random Forest achieved the highest overall accuracy (78%), with strong performance across all ripeness categories. It demonstrated the best balance between precision and recall, making it the most reliable classification model.
- XGBoost performed slightly worse (74% accuracy), though it maintained high recall for the Raw category. Suggests that XGBoost handles small sample classes better but may have limitations in distinguishing closely related ripeness stages.

- Gaussian Mixture Model (GMM) had the lowest accuracy (42%), indicating that it struggles to classify ripeness stages due to its unsupervised nature effectively. The model exhibits low precision and recall, particularly for the ready-to-eat category, indicating limited ability to capture complex spectral variations.

Why Results Differ Across Models?

- Random Forest and XGBoost leverage decision trees, which are well-suited for categorical classification with structured spectral data.
- GMM, being an unsupervised model, does not learn from labeled data, leading to poor classification performance. It lacks the feature-learning capability of supervised models like RF and XGBoost.
- Raw category had high recall across all models, likely due to its distinct spectral signature, while Harvestable and Ready-to-eat categories were harder to separate due to overlapping spectral patterns.

Best and Worst Model Performance

- Best Model Random Forest
  - Achieved the highest accuracy (78%) and strong precision-recall balance.
  - Outperformed others in correctly classifying Ready-to-eat avocados, which is crucial for harvest timing.
- Worst Model Gaussian Mixture Model (GMM)
  - Struggled to classify spectral data effectively due to its unsupervised nature.
  - Extremely low F1-score (0.24) for the Ready-to-eat category, making it unreliable for practical use.

The findings indicate that Random Forest is the most suitable model for avocado ripeness classification, offering a good balance between accuracy and robustness. XGBoost provides comparable performance but struggles slightly with precision. GMM is unsuitable for this classification task, reinforcing the importance of using supervised learning models for spectral-based classification.

These results highlight the effectiveness of machine learning in non-destructive fruit quality assessment, providing a foundation for further research in optimized spectral feature selection and deep learning-based classification.

Figure 10 presents the confusion matrices of the three machine learning models Random Forest, XGBoost, and Gaussian Mixture Model (GMM) used to classify avocado ripeness into Harvestable, Raw, and Ready-to-eat categories. The confusion matrices provide a detailed breakdown of actual versus predicted classifications, highlighting each model performance in distinguishing between ripeness stages.

Analysis of Confusion Matrices

1. Random Forest (Left Matrix)
   - Correctly classified most instances, particularly for the Ready-to-eat category (31 correct out of 39) and Harvestable category (35 correct out of 47).
   - Misclassification occurs mainly between Harvestable and Ready-to-eat avocados, with 12 misclassified as Ready-to-eat and 8 Ready-to-eat misclassified as Harvestable.
   - Perfect classification for Raw category (5/5) suggests distinct spectral features for unripe avocados.

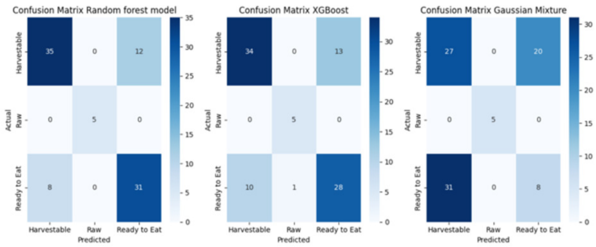

Fig.10: *Confusion Matrices for Random Forest, XGBoost, and Gaussian Mixture Models.*

2. XGBoost (Middle Matrix)
   - Slightly lower accuracy than Random Forest, with 34 correctly classified Harvestable and 28 correctly classified Ready-to-eat avocados.
   - Higher misclassification rate for Ready-to-eat avocados, with 10 incorrectly classified as Harvestable.
   - The Raw category remains ideally classified (5/5), similar to Random Forest.
3. Gaussian Mixture Model (Right Matrix)
   - The model shows significantly lower classification accuracy for ready-to-eat avocados, misclassifying 31 of 39 samples as harvestable.
   - Higher confusion between Harvestable and Ready-to-eat categories, indicating poor spectral feature separation.
   - The Raw category remains correctly classified (5/5), but this is due to its distinct spectral signature rather than model effectiveness.

Why the Differences?

- Random Forest and XGBoost are supervised learning models, allowing them to learn discriminative spectral features effectively.
- GMM is an unsupervised clustering model, meaning it does not leverage labeled data, leading to poor classification performance for closely related ripeness stages.
- The high confusion between Harvestable and Ready-to-eat in all models suggests overlapping spectral characteristics, making classification between these two stages more challenging.

Best and Worst Performing Models

- Best Model Random Forest – Achieves the most balanced classification with the highest correct classifications.
- Worst Model Gaussian Mixture Model (GMM) struggles significantly, particularly with Ready-to-eat avocados, demonstrating that an unsupervised approach is insufficient for this task.

The confusion matrix comparison confirms that Random Forest is the most effective model for ripeness classification, while XGBoost provides comparable but slightly lower performance. The Gaussian Mixture Model is unsuitable for this classification due to high misclassification rates. These results highlight the importance of supervised learning approaches in spectral-based fruit ripeness classification and suggest further refinement in feature selection to improve separation between closely related ripeness stages.

The confusion matrices for Random Forest, XGBoost, and Gaussian Mixture Model (GMM) show consistent misclassification between the Harvestable and Ready-to-eat categories, indicating a notable overlap in their spectral features. Random Forest misclassified 12 Harvestable avocados as Ready-to-eat and eight the other way, while XGBoost had a similar issue with 13 and 10 misclassifications. The bidirectional confusion indicates that the spectral signatures of these two ripeness stages are closely related, likely due to gradual biochemical transitions like softening and moisture reduction, which are hard to distinctly capture with a limited 18-band NIR sensor. The Raw category showed clear spectral separation, with perfect classification in the RF and XGBoost models, highlighting that the spectral differences between unripe and mature avocados are distinct and more straightforward for models to learn.

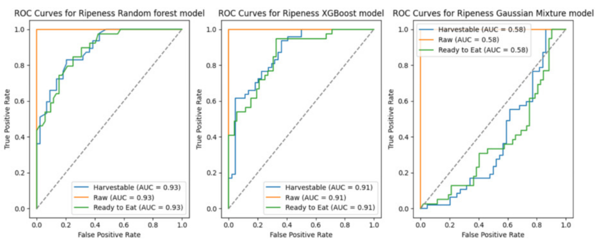

Fig.11: *ROC Curves for Random Forest, XGBoost, and Gaussian Mixture Models.*

Figure 11 presents the Receiver Operating Characteristic (ROC) curves for three machine learning models Random Forest, XGBoost, and Gaussian Mixture Model (GMM) used for avocado ripeness classification. The Area Under the Curve (AUC) values are provided for each ripeness category (Harvestable, Raw, and Ready-to-eat) to assess model performance in distinguishing between these classes.

Interpretation of ROC Curves and AUC Scores

1. Random Forest (Left Graph)
   - Achieved an AUC of 0.93 for all ripeness categories, demonstrating excellent classification ability.
   - The ROC curves remain close to the top-left corner, indicating a high actual positive rate and low false positive rate.
   - Suggests that Random Forest effectively differentiates between ripeness stages with firm decision boundaries.
2. XGBoost (Middle Graph)
   - AUC values of 0.91 across all categories, slightly lower than Random Forest, but still demonstrating strong classification capability.
   - The curves follow a similar trend to Random Forest, though minor deviations indicate slightly reduced sensitivity in distinguishing ripeness stages.
   - Overall, XGBoost is a highly effective model but slightly less precise than Random Forest.
3. Gaussian Mixture Model (Right Graph)
   - AUC values of 0.58 for all ripeness categories, indicating poor classification performance.
   - The ROC curves remain close to the diagonal (random classification line), meaning the model does not effectively separate the classes.
   - Confirms that GMM cannot accurately classify avocado ripeness due to its unsupervised nature, leading to significant overlap in spectral data.

Why the Performance Differences?

- Random Forest and XGBoost are supervised learning models that leverage decision trees and boosting techniques to extract relevant spectral features for classification.
- GMM, an unsupervised clustering model, lacks labeled training data, making it ineffective for classification tasks that require distinct decision boundaries.
- The high AUC values for Random Forest and XGBoost confirm their suitability for ripeness classification, while GMM poor performance suggests that clustering methods alone are insufficient for this task.

The ROC analysis highlights the superior performance of Random Forest and XGBoost, with high AUC values indicating strong classification capability. Random Forest performs best, followed closely by XGBoost, while GMM achieve only marginal classification accuracy. These results confirm that supervised machine learning models are essential for reliable spectral-based ripeness classification, ensuring more accurate predictions for harvest decision-making.

Figure 12 presents the Natural Language Processing (NLP)-generated interpretations of avocado ripeness predictions, providing a human-readable output based on the model classification results. The

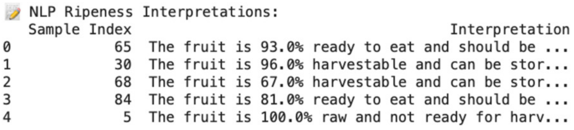

Fig.12: *NLP-Based Interpretation of Avocado Ripeness Predictions.*

system derives interpretations from the ripeness score by assigning probabilities to the raw, harvestable, and ready-to-eat stages and converting these numerical outputs into descriptive text. The inclusion of confidence percentages ensures transparency in prediction reliability, offering actionable insights for harvest decision-making.

The output format effectively bridges the gap between machine learning predictions and user comprehension, making it easier for farmers, agricultural experts, and stakeholders to understand the classification results. For example, the model classifies samples 0 and 3 as ready-to-eat with confidence scores of 93.0% and 81.0%, indicating suitability for near-term consumption. In contrast, the model labels samples 1 and 2 as harvestable with confidence levels of 96.0% and 67.0%, suggesting that they can be stored before consumption. Sample 5 receives a 100.0% raw classification, indicating that it is not ready for harvest and demonstrating the model's ability to deliver clear harvest recommendations.

This NLP-enhanced interpretation approach improves trust and usability in AI-driven fruit classification systems, ensuring that spectral-based ripeness assessments can be easily integrated into agricultural workflows. Future work could explore multi-language NLP integration, confidence interval refinement, and voice-based advisory systems, further enhancing accessibility and decision-making support in innovative farming applications.

5. CONCLUSIONS

This study expands on previous work regarding NIR-based avocado ripeness classification by broadening the task from binary (Raw/Harvestable) to three stages: Raw, Harvestable (Ripe), and Ready-to-Eat, utilizing the same low-cost hardware setup [26]. This research improves user interaction by using Natural Language Processing (NLP) to provide clear, human-readable descriptions of ripeness, moving away from the previous LED indicator system that required training to understand color codes. This enhances accessibility and simplifies the learning process for users. This enhancement tackles the usability issue observed in practical applications, especially for smallholder farmers.

Despite the limited dataset of 120 kg of Buccaneer avocados, the models achieved strong performance. Random Forest reached 78% accuracy with an AUC of 0.93, while XGBoost delivered comparable results. The preprocessing stage exposed the risk of excessive outlier removal, as it eliminated the entire raw category. To address this issue, the study adjusted IQR thresholds to preserve minority-class samples. Further error analysis showed spectral overlap between Harvestable and Ready-to-Eat classes, highlighting the necessity for improved feature extraction and a possible transition to ordinal or regression-based classification in future research. This study benchmarks three classical machine learning models; future work will extend the comparison to lightweight deep learning models (e.g., 1D-CNN and ResNet) and incorporate explainability methods to improve insight and fairness.

This study proposes a roadmap for practical deployment: quantizing trained models for offline inference on low-cost IoT devices, allowing operation without 5G or cloud access. The sensor hardware is still a prototype, utilizing an 18-band NIR sensor (410–940 nm) that needs additional calibration and casing improvements. The limitation of seasonal data and a single cultivar (Buccaneer) underscores the necessity for multi-season, multi-variety datasets to enhance generalizability, as noted by both reviewers. Future iterations will focus on farmer validation of NLP interpretations, incorporate Thai/local language support, and examine policy alignment with national postharvest standards and innovative farming frameworks.

This study demonstrates its feasibility of combining affordable NIR sensing, machine learning, and NLP to develop a precise, non-invasive method for classifying avocado ripeness. This is a significant step forward in making AI-driven precision agriculture more accessible, promoting digital inclusion, and helping small-scale farmers make better harvest decisions. The classification process, from spectral acquisition to NLP interpretation, takes less than 5 seconds per sample with a cloud-based system, showing it feasible for real-time use in agriculture.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Network Systems Laboratory Nara Institute of Science and Technology (NAIST), Japan, for providing research facilities, technical support, and valuable guidance throughout this study. The support and collaboration from the research team at NAIST have been instrumental in the successful completion of this work.

Furthermore, we extend our appreciation to the Faculty of Engineering, Rajamangala University of Technology Lanna (RMUTL), Thailand, for their continuous support and encouragement, which made this research possible.

AUTHOR CONTRIBUTIONS

Conceptualization, P.T., J.T., P.L., M.O. and K.Y.; methodology, P.T., P.L. and K.Y.; soft-ware, P.T., P.L. and K.Y.; validation, P.T., J.T., P.L., and K.Y.; formal analysis, P.T.,M.O. and P.L.; investigation, P.T., P.L. and K.Y.; data curation, P.T., P.L. and K.Y.; writing—original draft preparation, P.T.; writing—review and editing, P. T., P.L. and K.Y.; visualization, P.T., P.L., M.O. and K.Y.; supervision, J.T., P.L., M.O. and K.Y.; All authors have read and agreed to the published version of the manuscript.

References

[1] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto and A. Albarelli, "Fruit ripeness classification: A survey," *Artificial Intelligence in Agriculture*, vol. 7, pp. 44-57, 2023.

[2] M. Palumbo *et al.*, "Emerging Postharvest Technologies to Enhance the Shelf-Life of Fruit and Vegetables: An Overview," *Foods*, vol. 11, no. 23, p. foods11233925, 2022.

[3] C. Santos *et al.*, "Non-Destructive Measurement of the Internal Quality of Citrus Fruits Using a Portable NIR Device," *Journal of AOAC INTERNATIONAL*, vol. 104, no. 1, pp. 61-67, 2021.

[4] E. R. Arboleda, K. M. Parazo and C. M. Pareja, "Watermelon ripeness detector using near infrared spectroscopy," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 317-322, 2020.

[5] N. T. Anderson and K. B. Walsh, "Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation," *Journal of Near Infrared Spectroscopy*, vol. 30, no. 1, p. 09670335211057235, 2022.

[6] J. Walsh, A. Neupane, A. Koirala, M. Li and N. Anderson, "Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. II. The rise of convolutional neural networks," *Journal of Near Infrared Spectroscopy*, vol. 31, no. 3, p. 09670335231173140, 2023.

[7] P. Reddy *et al.*, "Near-Infrared Hyperspectral Imaging Pipelines for Pasture Seed Quality Evaluation: An Overview," *Sensors*, vol. 22, no. 5, p. s22051981, 2022.

[8] W. Zhang, L. C. Kasun, Q. J. Wang, Y. Zheng and Z. Lin, "A Review of Machine Learning for Near-Infrared Spectroscopy," *Sensors*, vol. 22, no. 24, p. s22249764, 2022.

[9] A. C. Agulheiro-Santos *et al.*, "Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy," *The Journal of the Science of Food and Agriculture*, vol. 102, no. 11, pp. 4866-4872, 2022.

[10] R. Pandiselvam *et al.*, "Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review," *Frontiers in Nutrition*, vol. 9, p. fnut.2022.973457, 2022.

[11] D. K. Bwambok *et al.*, "QCM Sensor Arrays, Electroanalytical Techniques and NIR Spectroscopy Coupled to Multivariate Analysis for Quality Assessment of Food Products, Raw Materials, Ingredients and Foodborne Pathogen Detection: Challenges and Breakthroughs," *Sensors*, vol. 20, no. 23, p. s20236982, 2020.

[12] Z. Du, W. Tian, M. Tilley, D. Wang, G. Zhang and Y. Li, "Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 21, no. 3, pp. 2956-3009, 2022.

[13] D. Wang *et al.*, "A Lightweight convolutional neural network for nicotine prediction in tobacco by near-infrared spectroscopy," *Frontiers in Plant Science*, vol. 14, p. fpls.2023.1138693, 2023.

[14] J.-R. Riba, R. Cantero, P. Riba-Mosoll and R. Puig, "Post-Consumer Textile Waste Classification through Near-Infrared Spectroscopy, Using an Advanced Deep Learning Approach," *Polymers*, vol. 14, no. 12, p. polym14122475, 2022.

[15] B. Benmouna *et al.*, "Convolutional Neural Networks for Estimating the Ripening State of Fuji Apples Using Visible and Near-Infrared Spectroscopy," *Food and Bioprocess Technology*, vol. 15, pp. 2226-2236, 2022.

[16] V. Gomes, A. Mendes-Ferreira and P. Melo-Pinto, "Application of Hyperspectral Imaging and Deep Learning for Robust Prediction of Sugar and pH Levels in Wine Grape Berries," *Sensors*, vol. 21, no. 10, p. s21103459, 2021.

[17] Y. Xu, W. Yang, X. Wu, Y Wang and J. Zhang, "ResNet Model Automatically Extracts and Identifies FT-NIR Features for Geographical Traceability of Polygonatum kingianum," *Foods*, vol. 11, no. 22, p. foods11223568, 2022.

[18] Z. Su *et al.*, "Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep Learning Approaches," *Frontiers in Plant Science*, vol. 12, p. fpls.2021.736334, 2021.

[19] P. Luekhong, P. Limkonchotiwat and T. Ruangrajitpakorn, "A Study on an Effect of Using Deep Learning in Thai-English Machine Translation Processes," *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Chiang Mai, Thailand, pp. 1-6, 2019.

[20] N. Ketui, K. Homjun, K. Poonyasiri, J. Deepinjai and P. Luekhong, "Item-based approach for online exam performance and its application," *2016 13th International Conference on Electrical Engineering/Electronics, Computer,*

Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, pp. 1-5, 2016.

[21] P. Mishra, D. N. Rutledge, J.-M. Roge, K. Wali and H. A. Khan, Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction," *Talanta*, vol. 229, p. 122303, 2021.

[22] P. Tipauksorn, P. Luekhong, J. Thongpron, U. Kamnarn, K. Yingkayun and A. Namin, "Stereo Vision-based Turn-Alignment Optimization for Wireless Power Transmission Positioning," *2023 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific)*, Chiang Mai, Thailand, pp. 1-8, 2023.

[23] N. Chen and M. Okada, "Toward 6G Internet of Things and the Convergence With RoF System," in *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8719-8733, 1 June1, 2021.

[24] A. Benelli, C. Cevoli and A. Fabbri, "In-field hyperspectral imaging: An overview on the ground-based applications in agriculture," *Journal of Agricultural Engineering*, vol. 51, no. 3, pp. 129-139, 2020.

[25] W. Ye *et al.*, "Application of Near-Infrared Spectroscopy and Hyperspectral Imaging Combined with Machine Learning Algorithms for Quality Inspection of Grape: A Review," *Foods*, vol. 12, no. 1, p. foods12010132, 2023.

[26] P. Tipauksorn, J. Thongpron, P. Luekhong, M. Okada and K. Yingkayun, "Low-Cost NIR and AI for Avocado Ripeness Classification: A Cloud-Integrated Approach," *2025 SICE Festival with Annual Conference (SICE FES)*, Chiang Mai, Thailand, pp. 1282-1287, 2025.

Panudech Tipauksorn received his B.Sc. in Industrial Education (Electrical Communication Engineering) from the Rajamangala Institute of Technology, Northern Campus, Thailand, in 1994. He completed his M.Sc. in Computer Engineering at King Mongkut's University of Technology Thonburi, Thailand, in 2006, and earned his D. Eng. in Electrical Engineering from Rajamangala University of Technology Lanna, Chiang Mai, Thailand, in 2025. He is currently an Assistant Professor in the Department of Computer Engineering at Rajamangala University of Technology Lanna. His research interests include computer networks, IoT, artificial intelligence, image processing, and cybersecurity.

Jutturit Thongpron is Asst. Prof. in the Department of Electrical Engineering, Rajamangala University of Technology Lanna (RMUTL), Chiang Mai, Thailand. He received his B.S.Tech.Ed. and M.S.Tech.Ed. in Electrical Engineering from King Mongkut's University of Technology North Bangkok (KMUTNB) in 1990 and 1994, respectively. His D.Eng. in Energy Technology is finished from King Mongkut's University of Technology Thonburi (KMUTT) in 2005. His main research interest is in photovoltaic (PV) characterizations, PV applications, smart grid technology, solar pumping, inductive wireless power transfer, wireless EV charging station, and, as well as electrical engineering education.

Prasert Luekhong works as a Lecturer at Rmutl, which is a College of Integrated Science and Technology, Rajamangala University of Technology Lanna. His B.Sc. in Electronic Industrial from Rajabhat University Chiang Rai, Thailand, in 2001 Master of Science in Internet and Information Technology from Naresuan University Phayao Campus, Thailand, in 2005 and Ph.D. Computer Science from Chiang Mai University, Thailand in 2016. He is currently an Assistant Professor with the College of Integrated Science and Technology, Rajamangala University of Technology Lanna, Chiang Mai, Thailand. His research interests include Cyber Security Big Data Natural Language Processing.

Minoru Okada received the B.E. degree in communications engineering from the University of Electro-Communications, Tokyo, Japan, in 1990, and the M.E. and Ph.D. degrees in communications engineering from Osaka University, Osaka, Japan, in 1992 and 1998, respectively. From 1993 to 2000, he served as an assistant professor. In 2000, he joined the Nara Institute of Science and Technology (NAIST), Nara, Japan, where he is currently a professor. His research interests include wireless communications and wireless power transfer. He is a member of the IEEE, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Institute of Image Information and Television Engineers (ITE).

Krisda Yingkayun received his B.Sc. in Industrial Education (Electrical Communication Engineering) from Rajamangala Institute of Technology, Northern Campus, Thailand, in 1991, and his master degree and doctoral degree in electrical engineering from faculty of engineering, Chiang Mai University, Thailand, in 1996 and 2013 respectively. He is currently an Assistant Professor with the Department of Electrical Engineering, Rajamangala University of Technology Lanna, Chiang Mai, Thailand. His research interests include Computer Programming, IoT, Embedded System, Machine Learning.