DETERMINATION OF OIL PALM FRUIT MATURITY USING A PORTABLE INSTRUMENT BASED ON UV-VIS-NIR SENSOR
/
PENENTUAN KEMATANGAN BUAH KELAPA SAWIT MENGGUNAKAN INSTRUMEN PORTABEL BERBASIS SENSOR UV-VIS-NIR

Yunisa Tri SUCI [1,2], I Wayan BUDIASTRA *[1,3], Y Aris PURWANTO[1], Slamet WIDODO[1]
[1] IPB University, Department of Mechanical and Biosystem Engineering / Indonesia;
[2] Jambi Agricultural Training Center, Ministry of Agriculture / Indonesia;
[3] IPB University, Centre for Research on Engineering Application in Tropical Agriculture / Indonesia ;
Tel: +622518623026; E-mail: wbudiastra@apps.ipb.ac.id
DOI: https://doi.org/10.35633/inmateh-76-49

***Keywords:** k-NN, multispectral sensor, oil palm fruit, precision agriculture, UV–Vis–NIR*

ABSTRACT
An accurate method for determining palm fruit maturity is needed by the palm oil industry to increase productivity and quality of crude palm oil. This study developed a portable instrument based on the AS7265x sensor (410–940 nm) to determine oil palm maturity accurately. Reflectance from 270 oil palm fruits at various maturity were measured, pre-processed, and classified into three maturity levels using PCA. PCA with SNV pre-treatment explained 97% of the variance. Classification was validated using SVM, RF, and KNN, with KNN achieving 100% accuracy. The instrument with SNV-PCA-KNN method has the potential to be used for oil palm fruit maturity classification.

ABSTRAK
Metode penentuan tingkat kematangan buah kelapa sawit dibutuhkan oleh industri sawit untuk meningkatkan produktivitas dan mutu minyak sawit kasar. Penelitian ini mengembangkan sebuah instrumen portabel berdasarkan sensor AS7265x (410-940 nm) untuk menentukan kematangan kelapa sawit. Reflektansi dari 270 buah kelapa sawit dengan beragam kematangan diukur, dilakukan pra-pengolahan, dan diklasifikasikan ke dalam tiga tingkat kematangan dengan menggunakan analisis komponen utama (PCA). PCA dengan pra-perlakuan SNV menjelaskan 97% dari varians. Klasifikasi divalidasi menggunakan SVM, RF, dan KNN, dengan KNN mencapai akurasi 100%. Instrumen dengan metoda SNV-PCA-KNN berpotensi diterapkan untuk klasifikasi kematangan buah kelapa sawit.

.

INTRODUCTION
Palm oil (*Elaeis guineensis*) is a major commodity crop, with global demand for palm oil continuing to increase due to its versatility and economic value. Timely and accurate harvesting is essential to maximize oil yield and quality *(Junior & Suharjito, 2023)*. Conventionally, harvesting is based on the maturity level of the fruit, which is determined by changes in fruit color and the detachment of fruit from the bunch. Manual classification is greatly affected by environmental factors like inconsistent lighting, rain, and shading, leading to misjudgments (*Mansour et al., 2022*). This time-consuming process relies on manual labor, limiting scalability and consistency in large plantations. Consequently, operational efficiency declines, which can negatively impact oil extraction rates (OER), resulting in lower yields, higher free fatty acid (FFA) content, reduced oil quality, and increased waste from suboptimal fruit (*Malyala, 2024*). These challenges highlight the limitations of manual methods for precise and large-scale harvesting especially in large plantations where variability in fruit maturity is high.

Recent studies have explored non-destructive to tackle the shortcomings of visual maturity assessment approaches such as image processing *(Saifullah et al., 2023),* electrical properties (Chin-Hashim et al., 2022), and near infrared (NIR)(Suci et al., 2024) *Budiastra et al., 2024*). Spectroscopy technologies have attracted growing interest due to their rapid detection capabilities and ability to simultaneously measure multiple constituents of oil palm fruit. Although these methods are promising, many rely on complex instrumentation and are limited in portability and field applicability, primarily due to the high costs of the necessary hardware and software.

Yunisa Tri SUCI, PhD.Stud.Agr. Eng; I Wayan Budiastra, Prof. Dr.Agr.Eng; Y Aris Purwanto, Prof.Dr.Agr.Eng; Slamet Widodo Assoc. Prof. Dr.Agr.Eng

Multichannel spectral sensors, particularly the AS7265x which covers the visible to near-infrared range (410–940 nm), offer a compact and cost-effective alternative for real-time spectral data acquisition. Despite its potential, the application of the AS7265x sensor in oil palm maturity classification remains underexplored. This sensor measures the absorption and reflection of light at particular wavelengths using various materials, including oil palm fruits. This sensor can provide precise information on the oil palm fruit's maturity by recognizing its spectral "fingerprints".

Researchers have used machine learning techniques and spectroscopic sensors to classify various agricultural products. A portable and low-cost spectrometric system using an AS7265x sensor and custom microcontroller was successfully used to classify coffee roasting levels using the Random Forest model *(Sagita et al., 2024)*, and to accurately determine the maturity of 'Ataulfo' mangoes *(Cano et al., 2024)*. Moreover, the integration of multichannel spectral sensors with machine learning models especially for on-site classification tasks has not been fully investigated in oil palm research. Prior work has focused on lab-based data acquisition or single-wavelength approaches, which do not reflect the complexities of dynamic outdoor environments. This study proposes an innovative solution by developing a portable, non-destructive instrument based on the AS7265x sensor, capable of real-time spectral acquisition, and coupling it with dimensionality reduction (PCA) and machine learning classifiers (SVM, RF, KNN) has the potential to be used for accurate maturity classification of oil palm fruitlets. The integration of spectral sensing and data-driven classification supports the advancement of precision agriculture, where timely and site-specific harvesting decisions are crucial for optimizing productivity and quality of crude palm oil. This approach not only addresses the limitations of existing manual and laboratory methods, but also contributes a novel, field-deployable solution that integrates multispectral sensing. This study aims to develop a portable instrument based on an AS7265X multichannel sensor and test its performance equipped with several machine learning algorithms for classifying the maturity level of oil palm fruits non-destructively.

MATERIALS AND METHODS

Portable instrument design based on multichannel spectral sensors

A portable instrument using multichannel sensors was developed to classify palm fruit maturity. The multichannel sensor used to build the palm oil detection system is the AS7265x sensor (9) of Austria Micro System, Styria, Austria. The sensor is explicitly composed of three TRIAD AS726x sensors, AS72651 (visible), AS72652 (UV) and AS72653 (IR) that detect absorbance at 18 wavelengths (410, 435, 460, 485, 510, 535, 560, 585, 610, 645, 680, 705, 730, 760, 810, 860, 900 and 940 nm). By default, UV LED emissions are recorded at six wavelengths from 410 to 535 nm, visible LED reflectance from 560 to 705 nm, and IR reflectance from 730 to 940 nm. This sensor allows for customizable LED configurations and data collection, enabling additional spectral data features. In this study, UV LED reflectance was also measured at 560 to 705 nm, known as fluorescence. This multispectral data acquisition sensor operates based on the principle of an optical reflective sensor. It features a light-emitting diode (LED) that emits light towards the palm kernel, which acts as the illuminated object. The light reflects off the palm oil fruit and is captured by a photodetector as the receiving sensor. The chemical composition of the oil palm influences the reflected light signal, resulting in a light wavelength signal that contains information about the chemical content of the palm fruit. The optoelectronic sensor then processes this information.

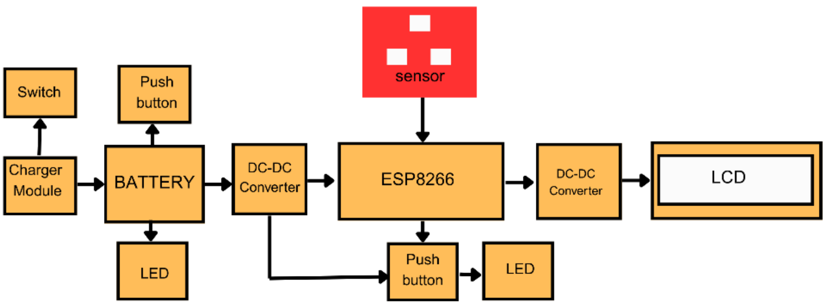


Fig. 1 - Component schematic of UV VIS NIR multichannel sensor acquisition equipment

The AS7265x multichannel sensor is connected to the ESP8266 via a jumper cable with resistors and integrated with a power supply (battery) for stable power supply. The electronic circuits are designed to be embedded in the PCB so that the connection of each component is stable. The data acquisition results will be sent to the cloud via Wi-Fi on the ESP8266 connected to a Wi-Fi router.

Data recorded in the form of date, time, and 24 channel spectrum data with symbols A, B, C, D, E, F, G, H, I, J, K, L, F1, F2, F3, F4, F5, and F6 sequentially.

The software used to control the microcontroller during data acquisition from the AS7265x sensor is Arduino IDE. The measurement results will be transferred in Microsoft Excel format, processed and analyzed using Unscrambler X for Principal component analysis (PCA), and classified using a machine learning algorithm. The component schematic of the portable instrument is summarized in Figure 1.

Sample preparation and spectra acquisition

The palm fruits used in the study were Tenera variety palm fruits from Cikabayan Plantation at IPB University, Bogor, Indonesia. The fruits were classified into 10 levels of harvesting age from young to old, starting at week 12 to 24 weeks of age, where at the age of 12 to 16 weeks, fruit collection was carried out once a week. The fruitless samples were taken from fresh fruit bunches from 3 different trees. A total of 9 fruits were taken from the center and apical parts of the FFBs, so the total number of loose fruit samples used was 270.

Spectra acquisition procedure is shown in Figure 2. The fruitlet was placed on a sample holder. A light beam from an LED exposed the fruitlet sample. Light reflected from the sample was captured by a photodetector with a 410-940 nm wavelength. The reflectance of the sample was measured by an instrument three times with different positions. The data acquisition results were sent to the cloud via Wi-Fi on the ESP8266 connected to a Wi-Fi router. The data spectral is recorded and stored in the cloud as date, time, and 24-channel spectrum data, which is then downloaded for further analysis.

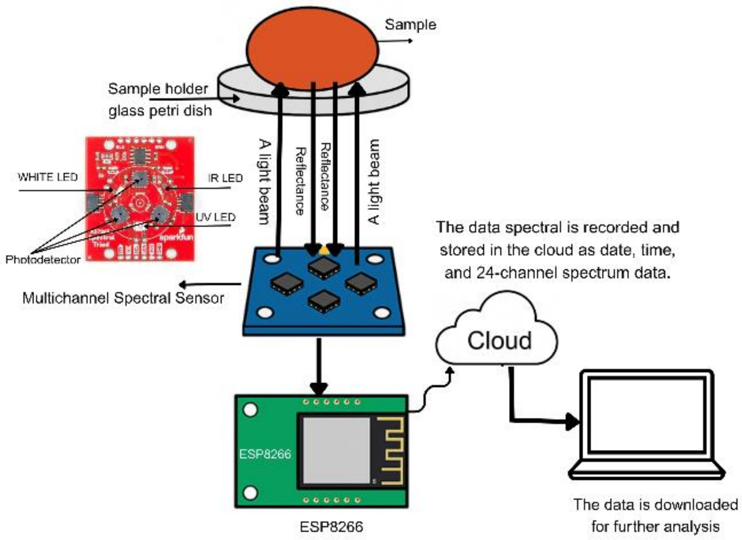


Fig. 2 - Spectra acquisition procedure

Classification of palm fruit maturity with machine learning

In this research, data from the AS7262X sensor is evaluated using PCA and machine-learning algorithms to classify oil palms according to their maturity level. This research utilizes three widely recognized machine-learning algorithms. RF, SVM, and KNN. The classification of oil palm fruit maturity, which includes 10 maturity levels, involves several stages: pre-treatment of spectroscopic data, dimensionality reduction using PCA, classification with machine learning models, and performance evaluation.

Pre-treatment of spectroscopic data

Raw reflectance data were pre-processed using three techniques: normalization, Standard Normal Variate (SNV), and Savitzky–Golay derivative filtering. These methods were applied to enhance signal quality, reduce noise, and correct for baseline shifts and light scattering effects, which are common in spectral data collected under field conditions. Spectral measurements obtained from multichannel sensors often exhibit variations caused by illumination changes, surface reflections, and environmental noise. The selected preprocessing techniques mitigate these issues by leveling the baseline, reducing scattering effects, and emphasizing subtle spectral features associated with chemical changes during fruit ripening. This preprocessing step is essential to improve feature differentiation and enhance the performance of subsequent classification models.

Principal component analysis (PCA)

After pre-processing, PCA was performed to reduce high-dimensional spectral data into uncorrelated variables (principal components) that capture significant variance related to fruit maturity. It aids in data compression and reveals clustering patterns among the 10 ripeness levels. Typically, the principal components explain 80% to 95% of the dataset's variability (*Michael et al., 2025*). PCA is effective for dimensionality reduction and identifying natural patterns in maturity data, serving as a foundation for advanced classification.

Classification using Machine Learning

To evaluate the effectiveness of the reduced features from PCA, three supervised machine learning algorithms SVM, RF, and KNN were implemented. These models were trained on the PCA-transformed data to classify the fruitlets into three consolidated maturity levels, which were derived from the original ten stages based on their PCA groupings. Based on the cluster pattern from PCA and referring to previous studies *(Eya'a et al., 2023)* the classification was consolidated into three maturity levels: unripe, ripe, and overripe. These categories are not only more physiologically representative but also more industrially relevant and applicable to support harvesting decisions. With this approach, the classification becomes more stable and accurate for application in portable sensor based non-destructive systems in the field. The machine learning classification process was conducted using Google Colab with Python 3.11.13 and the Scikit-learn package version 1.3.0. A total of 270 spectral samples from oil palm fruitlets were used, initially processed through group labelling into three maturity classes (unripe, ripe, overripe), followed by standardization to ensure uniform feature contribution, significant for distance-based algorithms such as KNN and SVM and class balancing to prevent bias toward dominant classes. The dataset was split into 70% training and 30% testing using stratified sampling to maintain class proportions and improve generalizability. Three classification models (SVM, RF, KNN) were trained on the training data, with hyperparameter optimization performed using GridSearchCV. This method conducts an exhaustive search over specified parameter grids via cross-validation to prevent overfitting and ensure optimal model configuration. Each model was then calibrated and validated on the test set and evaluated based on accuracy, precision, recall, and F1-score using both classification reports and confusion matrices. The final performance evaluation enabled the selection of the best-performing model by comparing all metrics, with the optimal model and its tuned parameters retained for potential deployment and future analysis.

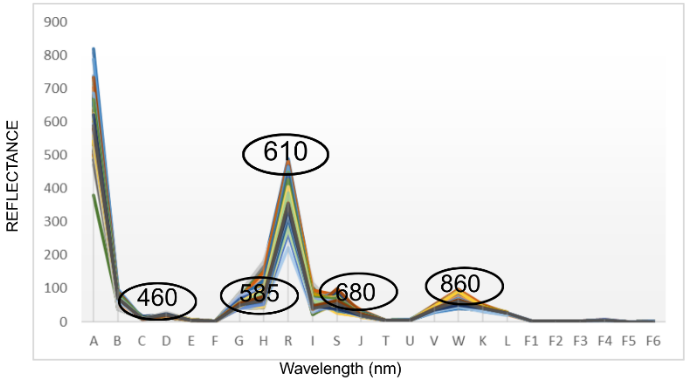
Performance metric

Classification metrics are crucial for evaluating the efficacy of models and directing advancements. They offer a way to assess how well a classification algorithm performs and give insight into how effectively a model is differentiating between classes. Metrics for evaluating classification models include accuracy, precision, recall, and F1 score *(Arifuddin et al., 2024).*

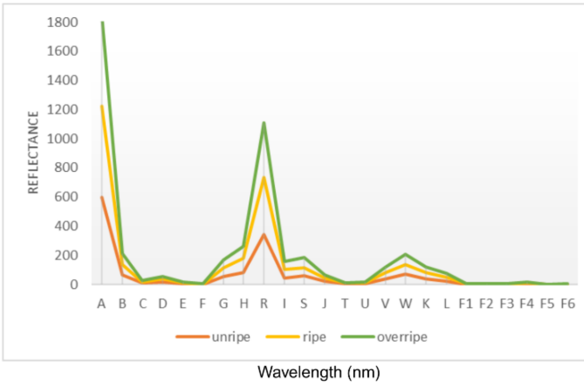
RESULTS

Characteristics of UV-VIS-NIR Spectra

The original spectra of oil palm fruit at various harvesting maturities within the UV-VIS-NIR wavelength range, obtained using a multichannel sensor, are illustrated in Figure 3.



a



b

Fig. 3 - Spectrum of oil palm fruits at different harvesting ages

a. Original spectrum b. averaged spectrum

The resulting spectra exhibit peaks and waveforms indicative of the absorption of chemical bonds by UV-VIS-NIR wavelengths. Prominent peaks are observed at wavelengths such as 460, 585, 610, 680, and 860 nm. Peaks within the UV-VIS region correspond to pigments in the fruit, including chlorophyll, carotenoids, xanthophylls, anthocyanins, and other phenolic compounds. As reflected in these pigments, fruit color is a critical quality attribute closely associated with the fruit's functional properties and developmental stages.

The spectra were averaged to discern the variations in peaks and wavelengths at every harvesting age of oil palm fruit (Fig. 3b). Although the peaks and wavelengths often lie within the same range, younger and older fruits can differ in values and peak positions. Carotenoids are present in the wavelength range of 430 nm to 510 nm, where these changes and variations are visible. Furthermore, alterations in the 535-560 nm range point to a fluctuation in the anthocyanin content, especially in the green color region. Spectral value variations are also seen in chlorophyll, which is detected in 680 and 720 nm regions. Changes in the NIR region occur in the 810-900 nm range, suggesting modifications in the O-H bond absorption.

Figures 3a and 3b show that classification based on the ten picking ages faces significant challenges due to overlapping spectra between groups, despite each age exhibiting variations in spectral values. This overlap causes the boundaries between groups to become less distinct, making the classification process more challenging. However, when the spectra were averaged and consolidated into three main age groups, unripe, ripe, and overripe, a more consistent and discrete shift in the spectra was observed. Classifying the data into three age groups for selection provides better separation and has the potential to enhance accuracy in the subsequent classification stage.

Classification of the sample using PCA

Separation and clustering were carried out to examine the differences in the characteristics of each picking age. Principal Component Analysis (PCA) was used to visualize palm fruits based on their characteristics. PCA is used to analyze VIS/NIR spectral data qualitatively. It synthesizes the data into principal components (PCs) with the highest spectral data variance.

PCA (Principal Component Analysis) can effectively represent the overall wavelength as input data for further processing and identify potential outliers. The original spectra of oil palm fruit are shown in Figure 4, which illustrates that three principal components (PCs) account for a total variance of 88%. Specifically, PC1 explains 54% of the variance, and PC2 contributes 34%. As for the classification based on the overall age of the input data, data processing still needs to be done so that it can be separated. Clustering with original data has formed clusters, but there is still visible overlap between PCs of each picking age. The original spectrum still contains background and noise in addition to information on the spectrum content for which pre-treatment is necessary.

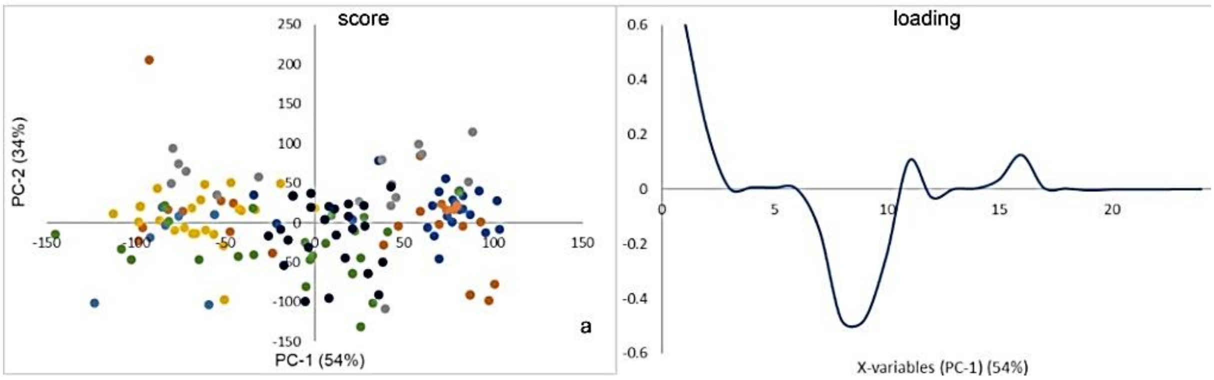


Fig. 4 - PCA spectra and loading plot

Figure 5 explains the difference in classification in the spectra averaged over several pre-treatments, while the total variance of each PC can be seen in Table 1. After the sample spectra data were averaged, it can be seen that the spectra data formed clusters at each picking age. The total variance explained by 2 PCs has exceeded 90% in both the original and pre-treatment data, but they still appear to overlap in the original spectra. PCA was performed on the processed spectral data, which had been subjected to various pre-treatment methods, including the original data, normalization, Standard Normal Variate (SNV), and Savitzky-Golay. The results show that the total variance explained by the two principal components (PC1 and PC2) varies depending on the pre-treatment method used. In the original data without pre-treatment, PC1 explained 71% of the variance, and PC2 20%, so the total variance captured by the two principal components was 91%.

However, when the data was processed using normalization and SNV methods, the value of PC1 increased significantly to 85%, while PC2 decreased to 11-12%, resulting in a total variance of 96-97%. This increase indicates that the data, which has been normalized or transformed with SNV, becomes more structured and exhibits a more dominant direction of the main variance in the first component. This indicates that the pre-treatment helps simplify the structure of the spectral data and improves the ability of PCA to separate important information.

Meanwhile, the Savitzky-Golay method yielded PC1 of 73% and PC2 of 17%, with a total variance of 90%. Although these results are promising, the explained variance value is still lower than that of normalization and SNV. Based on these results, it can be inferred that SNV is the most effective pre-treatment for improving PCA's representation ability of palm oil spectral data, so it is recommended for use in the subsequent classification stage.

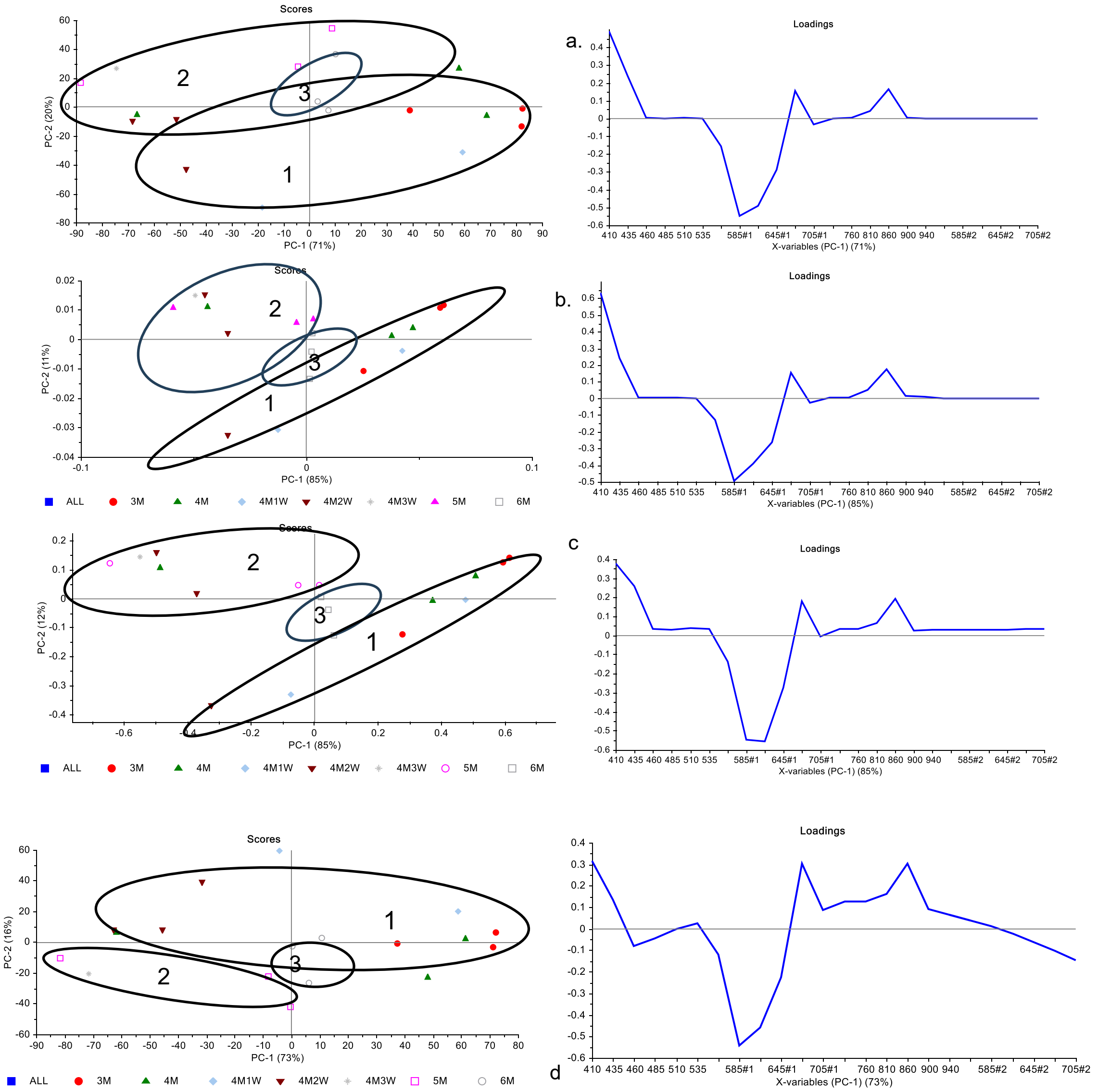


Fig. 5 - Sample average PCA spectrum and loading plot Sample average PCA spectrum and loading plot
(a) original (b) normalization (c) SNV (d) Savitzky Golay

Palm oil reflectance spectra frequently exhibit intensity changes due to sample surface, size, or thickness variances. SNV standardizes each spectrum independently by subtracting the mean and dividing it by the spectrum's standard deviation. This process removes the effects of baseline shift and scattering without altering spectrum's shape, maintaining the primary spectral information. As a result, variability unrelated to chemical components decreases, allowing PCA to capture more meaningful information in PC1. SNV enhances the dominance of Principal Component 1 (PC1) to 85% by making the differences in palm maturity more apparent than other disturbances, like measurement variation or noise. Unlike the Savitzky-Golay method, which reduces noise but ignores scattering, SNV effectively highlights fundamental inter-sample differences. Additionally, while normalization only scales data without addressing baseline disparities, SNV reveals actual inter-sample *(Rinnan et al., 2009; Zhu et al., 2025).*

Table 1

PC1 and PC2 values for each pre-treatment

No	Pre-treatment	PC1 (%)	PC2 (%)
1	Original	71	20
2	Normalization	85	11
3	SNV	85	12
4	Savitzky Golay	73	17

Classification of the sample using machine learning

To evaluate the accuracy of the clustering that has been formed from the PC classification using PCA and SNV, the oil palm spectra data is divided into unripe (spectra of 3 months, 4 months, 4 months 1 week and 4 months 2 week), ripe (4 months 3 weeks, 5 months, 5 months 1 week, 5 months 2 week), and overripe (5 months 3 week and 6 months) groups using SVM, RF, and KNN algorithms using phyton with the parameters used can be seen in Table 2. This classification is also based on standard harvesting practices in the field and refers to previous research *(Eya'a et al., 2023)*.

Table 2

Optimization parameters

Classification Method	Parameter	Range
SVM	Kernel Gamma C	Linear 0.7 10
KNN	n_neighbors weights metric p (distance)	5 uniform Minkowski 2
RF	n_estimators random_state	200 42

The selection of hyperparameters for each classification model (SVM, KNN, and RF) was carried out using GridSearchCV, a method that systematically explores combinations of predefined hyperparameter values through cross-validation. This approach ensures that the selected parameters yield optimal model performance on unseen data while minimizing the risk of overfitting. For SVM, a linear kernel was selected after comparing it with other kernels, as it yielded the most stable results while maintaining low computational complexity. The gamma parameter was set to 0.7, controlling the influence of individual data points, and the C value was set to 10, indicating a relatively substantial penalty for misclassification, which helps the model create a more accurate separating hyperplane.

In the case of KNN, the number of neighbors (n_neighbors) was optimized to 5, which balances bias and variance. The weights parameter was set to uniform, meaning that each neighbor has equal influence. The distance metric used was Minkowski with a p-value of 2, equivalent to the Euclidean distance, which is commonly used and effective for continuous data, such as spectra. For RF, the number of decision trees (n_estimators) was set to 200, providing a good balance between performance and computation time. The random_state was set to 42 to ensure the reproducibility of results. These parameters were selected based on their ability to reduce overfitting while capturing enough decision boundaries for accurate classification. GridSearchCV optimized model performance for classifying palm oil maturity levels using PCA-reduced spectral features.

SVM is well-suited for high-dimensional data, where the choice of kernel function plays a critical role in determining classification performance (*Pathak et al., 2022*). To enhance accuracy, spectral and spatial features were integrated using a joint kernel-based approach. In this study, meta-parameter optimization was performed using the regularization parameter C and the RBF kernel parameter G, with values set to C = 10, G = 0.7, and a linear kernel. This configuration yielded an accuracy of 0.80. The resulting confusion matrix presented in Figure 6.

Figure 7 shows one of the decision trees of a random forest. Random forests can reduce the impact of noise and outliers, resulting in more accurate and stable predictions. The random forest's ensemble nature allows it to average the predictions of each tree. Noise-containing predictions from individual trees can be mitigated when combined across the ensemble, resulting in more robust and stable predictions *(Raposo-neto et al., 2024)*.

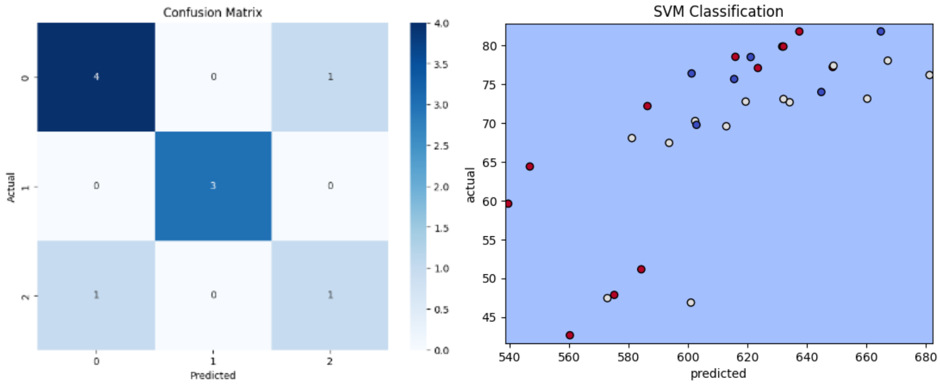


Fig. 6 - Confusion Matrix and decision boundary algorithm SVM classifier

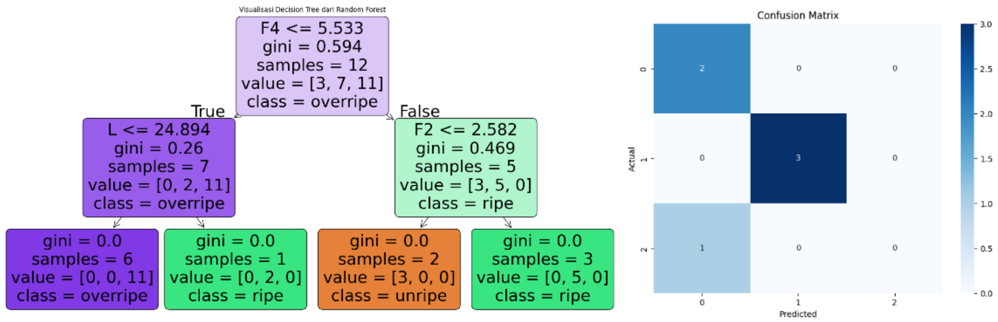


Fig. 7 - Decision tree of oil palm classification using random forest

Figure 8 shows the classification results using KNN. The KNN classification algorithm is one of the simplest machine learning algorithms with mature theory and wide application. The principle is to judge attributes based on the category of the k nearest points when predicting new values, which is simple and fast.

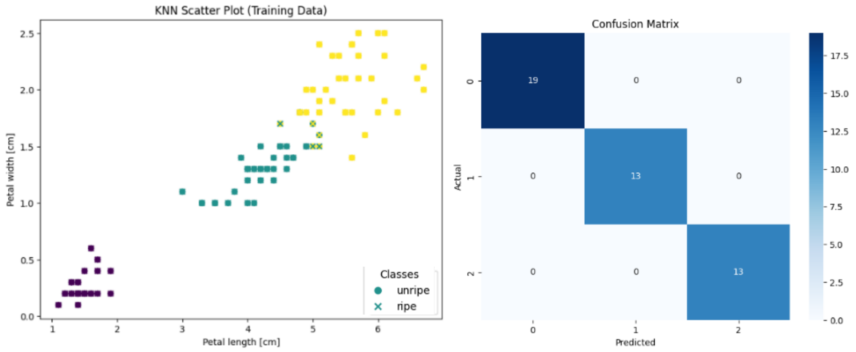


Fig. 8 - Scatter plot and confusion matrix of KNN classifier algorithm

Table 3 shows the classification results of the three models. The three algorithms used can classify well, but only KNN can classify with 100% accuracy, while SVM and Random Forest can only classify with 80% and 60% accuracy. KNN algorithm demonstrated 100% accuracy, indicating its effectiveness in distinguishing among the three classes. Several key factors contribute to KNN's superiority in this study. First, the relatively small number of features (18 wavelengths) and the limited sample size (270 spectra) enhance KNN's ability to learn data distribution efficiently, minimizing the risk of high computational costs. Second, the initial clustering performed using Principal Component Analysis (PCA) likely led to clear class separations in the feature space. This clarity allows distance-based algorithms like KNN to classify samples based on their nearest neighbors. In this study, a value of k=5 was utilized employing the Minkowski metric (p=2, which is equivalent to Euclidean distance) and applied uniform weighting. This approach ensures that classification is determined by averaging the five closest samples without giving additional weight to nearer neighbors. When the class distribution in the feature space is well-separated, the KNN model tends to achieve very high accuracy. KNN is known for its high computational complexity. It requires calculating the Euclidean distance between the input feature and every feature in the database.

While it does not involve a training phase, it is computationally intensive during the classification phase (*Eya'a et al., 2023*), *Saifullah et al., (2023*) utilized K-Nearest Neighbors (KNN) with RGB and Lab color extraction for automatic oil palm maturity classification, achieving 95.4% and 97.4% accuracy, respectively.

Table 3

Classification results

Parameter	SVM	RF	KNN
Accuracy	0.80	0.60	1.00
Recall	0.77	0.60	1.00
Precision	0.77	0.69	1.00
F_1 score	0.77	0.58	1.00

The Support Vector Machine (SVM) model achieved an accuracy of 80%, which is lower than the K-nearest neighbors (KNN) model but still better than the Random Forest (RF) model. SVM operates by constructing an optimal hyperplane to separate different classes in the feature space. This study used a linear kernel with parameters of gamma = 0.7 and C = 10. Linear kernels separate classes using straight lines, which can be a limitation when the data distribution is not linear. SVM is limited in speed and size during both the training and testing phases of the algorithm, and it is also limited in speed concerning the selection of the kernel function parameters *(Kumar et al., 2024)*. It is likely that the results from Principal Component Analysis (PCA) still exhibit non-linear patterns that cannot be fully captured by a linear-based model, leading to some misclassification near the edges of class separations.

Meanwhile, with its 60% accuracy, Random Forest demonstrated the lowest performance in capturing data distribution patterns after PCA. This study configured RF with 200 decision trees (n_ estimators = 200) and random_ state = 42. Despite RF's theoretical ability to handle complex data, the low results suggest that the PCA data structure may be too simple for a decision tree-based approach. The complexity of the data is a challenge that needs to be addressed, as it could be a contributing factor to the model's performance. Another possibility is overfitting on some subsets of the data, so RF fails to generalize well to the entire dataset. Random Forest (RF) can be overly sensitive to minor variations in the training set, sometimes leading to instability and a tendency to overfit. In contrast, Support Vector Machines (SVM) and RF are generally resilient to noise and excessive training, indicating their effectiveness in handling unbalanced data *(Kausik et al., 2025)*.

CONCLUSIONS

A portable instrument based on the AS7265x multichannel spectral sensor integrated with machine learning algorithms was developed and tested for oil palm fruit maturity classification. PCA analysis of SNV-pre-treated reflectance spectra revealed that the first two principal components accounted for 97% of the variance, resulting in a clear separation among the three maturity levels. Among the evaluated models, KNN achieved the highest classification accuracy of 100%. The system has the potential to be used for oil palm fruit maturity classification and to reduce human error in visual maturity assessment. However, the system has also some limitations such as classification is still processed out of the instrument, and the number of samples used is relatively limited. Further research is needed to ensure this system can be used as an accurate and real-time method for oil palm fruits classification to support precision agriculture such as integration of the trained model into the instrument as well as system validation in large number of oil palm fruit samples.

ACKNOWLEDGEMENT

The authors are grateful to Directorate General of Higher Education Research, and Technology under Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for the Research Grant (PDD) of Program Implementation Contract for the Year 2024: 027/E5/PG.02.00.PL/2024 and the Agency of Agricultural Extension and Human Resources Development, Ministry of Agriculture, Indonesia for supporting this research.

REFERENCES

[1] Arifuddin, A., Buana, G. S., Vinarti, R. A., & Djunaidy, A. (2024). Performance Comparison of Decision Tree and Support Vector Machine Algorithms for Heart Failure Prediction. *Procedia Computer Science*, *234*, 628–636. https://doi.org/10.1016/j.procs.2024.03.048

[2] Budiastra, W., Marjan, S., Adiarifia, N., Novianty, I., & Suci, Y. T. (2024). Non-Destructive Prediction of Oil and Free Fatty Acid of Oil Palm Fruitlets Using Near-Infrared Spectroscopy and Hybrid Calibration Method. *INMATEH - Agricultural Engineering*, *73*(2), 461–472. https://doi.org/10.35633/inmateh-73-39

[3] Chin-Hashim, N. F., Khaled, A. Y., Jamaludin, D., & Abd Aziz, S. (2022). Electrical Impedance Spectroscopy for Moisture and Oil Content Prediction in Oil Palm (Elaeis guineensis Jacq.) Fruitlets. *Plants*, *11*(23). https://doi.org/10.3390/plants11233373

[4] Eya'a, N. C., Nsimi, M. A., Ndele, P. H., Ngalle, B. H., Molo, T., Mbo, N. L. F., Fouman, A., Likeng, L.-G. B., Ngando, E. G. F., & Bell, J. M. (2023). A Review of Main Factors Involved in the Maturation of Oil Palm (Elaeis guineensis Jacq.) Fruit Bunches. *American Journal of Plant Sciences*, *14*(07), 727–740. https://doi.org/10.4236/ajps.2023.147049

[5] Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information (Switzerland)*, *15*(4), 235. https://doi.org/10.3390/info15040235

[6] Junior, F. A., & Suharjito. (2023). Video based oil palm ripeness detection model using deep learning. *Heliyon*, *9*(1), e13036. https://doi.org/10.1016/j.heliyon.2023.e13036

[7] Kausik, A. K., Rashid, A. Bin, Baki, R. F., & Jannat Maktum, M. M. (2025). Machine learning algorithms for manufacturing quality assurance: A systematic review of performance metrics and applications. *Array*, *26*(2025), 100393 Contents. https://doi.org/10.1016/j.array.2025.100393

[8] Kumar, A., Gaur, N., & Nanthaamornphong, A. (2024). Machine learning RNNs, SVM and NN Algorithm for Massive-MIMO-OTFS 6G Waveform with Rician and Rayleigh channel. *Egyptian Informatics Journal*, *27*(March), 100531. https://doi.org/10.1016/j.eij.2024.100531

[9] Malyala, R. (2024). Development of a Convolutional Neural Network Model for Automated Ripeness Classification of Palm Oil Fresh Fruit Bunches. *International Journal of Innovative Science and Research Technology (IJISRT)*, *9*(11), 1040–1046. https://doi.org/10.38124/ijisrt/ijisrt24nov762

[10] Mansour, M. Y. M. A., Dambul, K. D., & Choo, K. Y. (2022). Object Detection Algorithms for Ripeness Classification of Oil Palm Fresh Fruit Bunch. *International Journal of Technology*, *13*(6), 1326–1335. https://doi.org/10.14716/ijtech.v13i6.5932

[11] Michael, N. A., Ibrahim, M. A. Al, Scheibe, C., & Craigie, N. (2025). Best practices of utilizing Principal Component Analysis (PCA) in chemostratigraphic studies Data ro o f. *Applied Geochemistry*, 1–12. https://doi.org/10.1016/j.apgeochem.2025.106355

[12] Raposo-neto, J. J., Kowalski-neto, E., Luiz, W. B., Fonseca, E. A., Karla, A., Logrado, C., Singh, M. N., Martin, F. L., Vassallo, P. F., Cristina, L., Campos, G., & Barauna, V. G. (2024). Near-Infrared Spectroscopy with Supervised Machine Learning as a Screening Tool for Neutropenia. *Personalized Medicine*, *14*(9), 1–13. https://doi.org/https:// doi.org/10.3390/jpm14010009 Academic

[13] Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, *28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

[14] Saifullah, S., Prasetyo, D. B., Indahyani, Drezewski, R., & Dwiyanto, F. A. (2023). Palm Oil Maturity Classification Using K-Nearest Neighbors Based on RGB and L*a*b Color Extraction. *Procedia Computer Science*, *225*, 3011–3020. https://doi.org/10.1016/j.procs.2023.10.294

[15] Suci, Y. T., Budiastra, I. W., Purwanto, Y. A., Widodo, S., & Novianty, I. (2024). Wavelengths Selection Method in Near Infrared Spectra Using Multivariate Analysis for Nondestructive Determination of Oil Content in Palm Oil Fruits. *IOP Conference Series: Earth and Environmental Science*, *1386*(1). https://doi.org/10.1088/1755-1315/1386/1/012009

[16] Zhu, J., Zhu, X., Yan, B., Ren, F., Chen, B., Han, Z., Yao, X., He, S., & Liu, H. (2025). Evaluation and categorization of various pea cultivars utilizing near-infrared spectroscopy in conjunction with multivariate statistical techniques. *Food Chemistry*, *474*(January), 1–15. https://doi.org/10.1016/j.foodchem.2025.143268