IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 4, August 2025, pp. 2955~2963
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i4.pp2955-2963

Enhancing crude palm oil quality detection using machine learning techniques

Novianti Puspitasari[1], Ummul Hairah[1], Vina Zahrotun Kamila[2], Hamdani Hamdani[1], Anindita Septiarini[1], Amin Padmo Azam Masa[2]
[1]Department of Informatics, Faculty of Engineering, Mulawarman University, Samarinda, Indonesia
[2]Department of Information System, Faculty of Engineering, Mulawarman University, Samarinda, Indonesia

Article Info

Article history:

Received Sep 8, 2024
Revised May 10, 2025
Accepted Jun 8, 2025

Keywords:

Crude palm oil
Data acquisition
Machine learning
Pre-processing
Quality evaluation

ABSTRACT

Indonesia, a leading nation in the palm oil industry, experienced a significant increase of 15.62% in crude palm oil (CPO) exports in 2020, effectively meeting the global need for vegetable oil and fat. Therefore, the subjective assessment of CPO quality, influenced by differences in human evaluations, may lead to inconsistencies, necessitating the adoption of machine learning methods. There are several categories of CPO, such as bad and excellent. Machine learning can determine the quality of CPO itself. This study utilizes two distinct categories to measure the quality of CPO. CPO quality data is collected and processed into pre-processing data, in classifying using several methods such as artificial neural network (ANN), k-nearest neighbor (KNN), support vector machine (SVM), decision tree (DT), naïve Bayes (NB), and C.45 using the cross-validation evaluation parameter. The best results are obtained by C.45 and DT with an accuracy of 99.98%.

Corresponding Author:

Novianti Puspitasari
Department of Informatics, Faculty of Engineering, Mulawarman University
Sambaliung St. no. 9, Samarinda, Indonesia
Email: novipuspitasari@unmul.ac.id.

1. INTRODUCTION

Indonesia, an agricultural nation, is a major palm oil producer due to its large land area and low production cost. The country's palm oil production rose 14% in August 2019, with a 33.88% increase in output from people's plantations and 58.56% growth from private large plantations. However, the state plantation (SP) has a sluggish output rate of 7.55%. The country's palm oil industry contributes to economic prosperity [1]. Indonesia's crude palm oil (CPO) exports increased by 15.62% in 2020 to 26.47 million tons, meeting the growing global demand for vegetable oils and fats [2]. Superior CPO is produced using mature oil palm fruits, classified as unripe for raw, ripe for harvestable, and overripe for entirely ripe. This results in high oil extraction efficiency and low free fatty acids (FFA). The subjective nature of CPO quality assessment due to individual appraisal differences can lead to inconsistencies, potentially lowering accuracy and objectivity, thus necessitating machine learning methods for CPO quality assessment.

Numerous researchers have thoroughly investigated the utilization of computerized technology in agriculture. This task includes IoT [3], remote sensing data [4], land suitability [5], identifying disease on fruit [6], leaves [7]–[9] and stem [10], segmenting fruit [11], [12], estimating the mass of fruit [13], [14], and fruit maturity [15]–[17]. In classification, several methods have been applied in several machine learning such as artificial neural network (ANN) [18], [19], support vector machine (SVM) [20]–[22], random forest (RF) and gradient boosting [3], [23], decision tree (DT) [24], [25], and k-nearest neighbor (KNN) [26], [27].

Several studies related to oil palm objects were developed by implementing machine learning approaches. Three machine learning algorithms were employed: multilayer perceptron, support vector regression (SVR), and linear regression to forecast CPO production. The SVR method surpassed the other two in prediction accuracy, with a positive predictive accuracy of 0.694, mean squared error (MSE) of 1146.054, mean absolute percentage error (MAPE) of 47.485, and mean absolute deviation (MAD) of 22.333 [1]. Three machine learning models were analyzed using historical data from 2020 to 2023 to predict CPO prices. The RF method showed superior performance; in the 90:10 scenario, RF outperformed linear and logistic regression, yielding smaller MSE (43948.56), MAE (80.37), and RMSE (209.64). Similarly, in the 80:20 scenario, the RF had smaller MSE (137787.61), MAE (106.38), and RMSE (371.20). In the 70:30 scenario, the RF showed smaller MSE (107582.32), MAE (104.13), and RMSE (328) [28].

The long short-term memory (LSTM) and extreme gradient boosting (XGBoost) models were evaluated by performing hyperparameter tuning optimization using multivariate data to find the most optimal model in forecasting CPO production with the lowest error rate. The results showed that the LSTM model produced better prediction results after hyperparameter tuning with an accuracy rate of 93.7% and RMSE of 21.04. The XGBoost model also experienced improved performance after tuning, with an RMSE of 22.17 and an accuracy rate of 92.8% [29]. The machine learning framework for oil palm breeding was applied with a primary emphasis on phenotypic data rather than genetic variables. The proposed model incorporated multiple methodologies, including SVM, ANN, and RF, which exhibited exceptional precision in forecasting variables such as oil production and bunch amount. Additionally, the framework facilitated the identification of high-yielding, stress-tolerant oil palm cultivars for sustainable agricultural production [30]. This research aims to establish a classification method for evaluating the grade of CPO, categorized into superb and substandard. The method entails separating unprocessed data into components for efficient analysis, which is achieved through data purification, transformation, normalization, and resampling using machine learning algorithms.

2. METHOD

This study aims to calculate the quality of CPO by gathering information from PT. Telen Prima Sawit Muara Bengkal. The data collected from January 2019 to August 2023 includes other characteristics, including FFA, moisture content, dirt content, and the deterioration of bleachability index (DOBI). The classification process is carried out through cleaning, transformation, normalization, splitting, and resampling. The methodology of CPO quality evaluation is displayed in Figure 1.

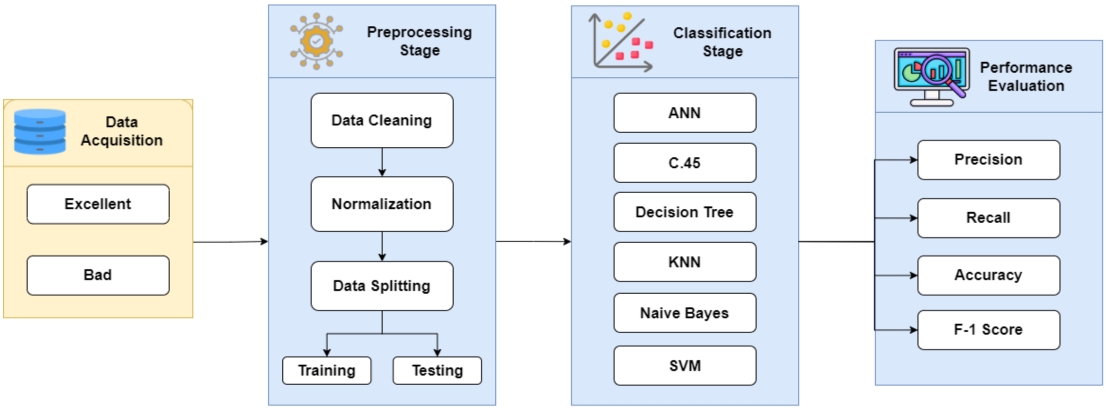

Figure 1. Proposed methodology of CPO quality evaluation

2.1. Dataset

This study utilizes 6,790 high-quality CPO primary data provided by PT. Telen Prima Sawit Muara Bengkal mill, spanning from January 2019 to August 2023. The data comprises variables such as FFA, moisture, dirt, and DOBI. Indonesian Palm Oil mill establishes CPO quality standards. Machine learning was employed in the study to analyze the performance of the CPO dataset. Table 1 shows the CPO quality data variable.

Direct field observations were used to collect the raw data on CPO quality. Subsequently, the data was classified in accordance with the Indonesian Palm Oil mill's exhaustive quality standards. Table 2 shows examples of CPO quality data. The dataset consists of two classes: excellent and bad. The excellent class of CPO quality data has an FFA range of 2.5-3.5%, a maximum moisture content of 0.15%, a maximum impurity level of 0.02%, and a minimum FFA content of 2.5%. The bad class has levels over the specified range. The "#N/A" label indicates the absence of variable or missing values. The CPO quality classification procedure excludes data with this label during data cleaning.

Table 1. CPO quality data variable

Variable	Information	CPO quality standards for Indonesian Palm Oil mill
FFA	Percentage of FFA in CPO.	2.5–3.5%
Moisture	Percentage of water content in CPO.	0.15% max
Dirt	Percentage of dirt and foreign material in CPO.	0.02% max
DOBI	Index that measures the ability of CPO to undergo the bleaching process effectively	2.5% min

Table 2. The example of CPO data quality

Variable				Label
FFA	Moisture	Dirt	DOBI	
2.99	0.15	0.02	-	Bad
-	-	-	-	#N/A
-	-	-	-	#N/A
-	-	-	-	#N/A
4.88	0.29	0.02	2.34	Bad
4.30	0.17	0.02	2.62	Bad
4.29	0.17	0.02	2.62	Bad
:	:	:	:	:
2.65	0.15	0.02	2.97	Excellent
2.64	0.15	0.02	2.97	Excellent
2.6	0.15	0.02	2.98	Excellent

2.2. Data pre-processing

The preprocessing stage is the initial step in data processing, which includes various procedures to prepare raw data before it is used in a machine learning model. This process involves several key steps, including normalization, data transformation, categorical variable encoding, and feature scaling. Each of these steps aims to enhance data quality and facilitate the model's ability to identify relevant patterns, thereby significantly improving accuracy. The series of stages in preprocessing is further explained through descriptive points that detail each step. Starting from initial cleaning to converting data into a form suitable for model training, all these stages are carried out systematically. This stage not only prepares the data technically but also plays a crucial role in enhancing the overall performance of the machine learning model.

2.2.1. Data cleaning

During the data cleaning stage, the quality of CPO data is analyzed to identify issues such as missing values, outliers, and incomplete data. This process is crucial to ensure that the analyzed data is accurate and consistent. The data cleaning process includes filling in missing values, correcting incomplete or inconsistent data, and removing irrelevant data. Outliers are also identified using certain statistical approaches to determine whether they reflect errors or are part of the natural variation of the data. Table 3 presents an example of CPO quality data that has undergone the cleaning process. This table shows how the initially irregular data has been selected and refined so that it can be used effectively in the further analysis process. The cleaned data can then be further analyzed.

2.2.2. Data transformation

Data transformation is a follow-up step to data cleaning, aiming to modify the data format to make it more suitable for use in machine learning modeling. In this context, the label encoding method is used to transform the CPO quality data labels that have undergone the cleaning stage. This process is important because many machine learning algorithms can only work with numeric data; therefore, category labels such as "high," "medium," or "low" need to be converted into a numeric form. Label encoding changes each category in the CPO quality data into a unique numeric value. In this study, the number "1" represents the "very good" class. Conversely, the number "0" means the "bad" class. This transformation not only simplifies the data structure but also increases the efficiency of the model training process. Table 4 presents an example of CPO quality data encoded using the label encoding method. The table displays the final results of the transformation process, where qualitative labels have been successfully converted into numeric representations, ready for use in classification or prediction processes. The right data transformation has a significant impact on the quality of the model and the accuracy of the results obtained.

Table 3. CPO quality data cleaning results

Variable				Label
FFA	Moisture	Dirt	DOBI	
3.03	0.15	0.02	2.91	Excellent
2.65	0.15	0.02	2.97	Excellent
:	:	:	:	:
4.4	0.14	0.02	2.77	Bad
4.27	0.16	0.02	2.73	Bad

Table 4. CPO quality data transformation results

Variable				Label
FFA	Moisture	Dirt	DOBI	
2.94	0.15	0.02	2.91	0
3.03	0.15	0.02	2.91	1
3.1	0.16	0.02	2.93	0
3.17	0.16	0.02	2.91	0
:	:	:	:	:
4.4	0.14	0.02	2.77	0
4.27	0.16	0.02	2.73	0
4.28	0.16	0.02	2.73	0
4.24	0.16	0.02	2.73	0

2.2.3. Data normalization

Data normalization is a crucial step in the preprocessing stage, particularly when working with numerical data, such as CPO quality data. At this stage, the variables in the dataset that have undergone the cleaning process will be normalized to a uniform scale. One method used is min-max normalization, which aims to set the values of the variables in a certain range, usually between 0 and 1. Normalization is performed to prevent scale imbalances between variables, which can lead machine learning algorithms to be more biased towards variables with larger values. By creating a proportional scale, this process enables fair comparisons between features and enhances the performance of scale-sensitive algorithms, such as KNN and SVM. This process does not alter the distribution of the data but rather rearranges the scale of the numerical values to be parallel. Table 5 presents the minimum and maximum values of each variable in the CPO quality dataset. These values are the main reference in calculating min-max normalization. Meanwhile, Table 6 presents an example of the results obtained by normalizing the data using (1).

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Table 5. CPO quality data transformation results

Variable	Min	Max
FFA	2.25	5.38
Moisture	0.14	0.17
Dirt	0.02	0.02
DOBI	2.72	2.99

Table 6. CPO quality data normalization results

Variable				Label
FFA	Moisture	Dirt	DOBI	
0.249201	0.333333	0	0.703704	1
0.271565	0.666667	0	0.777778	0
0.29393	0.666667	0	0.703704	0
0.268371	0.666667	0	0.777778	0
⋮	⋮	⋮	⋮	⋮
0.124601	0.333333	0	0.925926	1
0.111821	0.333333	0	0.962963	1
0.686901	0	0	0.185185	0

2.3. Classification

Six machine learning classification techniques were used to predict the quality of CPO after data was ready for modeling. These techniques include SVM, naïve Bayes (NB), KNN, DT, and ANN [18]. SVM categorizes data by separating it with a hyperplane and categorizes items into specific classes. The NB algorithm computes probabilistic results by combining data values, assuming that all characteristics are independent. The KNN algorithm utilizes a hierarchical data structure to calculate the distance between a certain location and the points in the designated training dataset. A hierarchical framework, an algorithmic DT, is used to forecast diabetes mellitus, focusing particularly on geographical regions and characteristics within a defined domain. Disease data categorization in machine learning can be achieved by utilizing various DT algorithms like iterative dichotomizer 3 (ID3), J48, C4.5, C5, chi-squared automatic interaction detection (CHAID), and classification and regression trees (CART). Due to its enhanced capabilities, the C4.5 approach was employed to maximize the performance analysis of diabetic data.

2.4. Performance evaluation

The evaluation parameters used to measure the performance of the machine learning methods are accuracy, precision, recall, and F1-score. The performance of the methods for detecting the quality of CPO. Therefore, the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) have to be calculated. TP is the data number of an excellent class classified as excellent, while TN is the data number of a bad class classified as bad. FP is the data number of an excellent class classified as a bad class, while FN is a bad class classified as an excellent class. Table 7 shows the confusion matrix for CPO quality detection. The evaluation parameters are defined in (2)–(5) [31].

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \tag{2}$$

$$Recall = \frac{(TP)}{(TP+FN)} \tag{3}$$

$$Precision = \frac{(TP)}{(TP+FP)} \tag{4}$$

$$F1-score = \frac{2\times(Precision \times Recall)}{(Precision+Recall)} \tag{5}$$

Table 7. The confusion matrix for CPO quality detection

	Predicted bad (0)	Predicted excellent (1)
Actual bad (0)	TN	FP
Actual excellent (1)	FN	TP

3. RESULTS AND DISCUSSION

The study investigates the potential of machine learning in enhancing CPO quality evaluation. Six popular algorithms, including ANN, C4.5, DT, KNN, NB, and SVM, were evaluated using a rigorous framework. The cross-validation method with the k-fold value of 3, 5, and 10 ensured a strong performance evaluation, reducing potential biases. Four metrics: precision, recall, accuracy, and F1-score were used to measure the effectiveness of each classifier in predicting CPO quality. This approach provided a detailed assessment of their strengths and weaknesses in the field. Table 8 presents the results of the classification of each classifier along with the k-fold value.

Table 8. The detection result of CPO quality using several machine learnings

Classifier	K-fold 3 (%)				K-fold 5 (%)				K-fold 10 (%)			
	Precision	Recall	Accuracy	F1-score	Precision	Recall	Accuracy	F1-score	Precision	Recall	Accuracy	F1-score
ANN	80	82.5	82.46	80.9	78.7	83.6	83.58	79.4	78.2	83.7	83.74	78.5
C.45	100	100	99.98	100	100	100	99.98	100	100	100	99.98	100
DT	100	100	99.98	100	100	100	99.98	100	100	100	99.98	100
KNN	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.6	99.6	99.6	99.6
NB	89.9	74.4	74.39	77.8	90	75.1	75.08	78.4	90	75.2	75.19	78.5
SVM	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1

The k-fold value of 3 was employed to observe early performance trends and make algorithm-specific observations. The C4.5 and DT algorithms demonstrated exceptional performance, achieving ideal scores of 100% in all evaluation metrics. The KNN algorithm demonstrated exceptional performance, achieving 99.4% accuracy, underscoring the efficacy of local similarity in classification. Although ANN attained 80% precision, recall, accuracy, and F1-score, its performance was inferior to that of other models. NB demonstrated a 77.8% F1 score, 74.39% accuracy, 74.4% recall, and 89.9% precision; however, its performance was subpar. SVM's capacity to manage intricate data distributions was illustrated by its consistent achievement of over 84% accuracy.

The k-fold value of 5 was applied to C4.5, and DTs are machine learning algorithms that consistently outperform others in assessing CPO quality. The significance of parameter optimization was underscored by the 99.5% accuracy of KNN, which had a k-value of 5. Despite a minor improvement, ANN could not surpass the top-performing algorithms, indicating that the constraints of capturing delicate data patterns may not be resolved by increasing the amount of training data. NB, which exhibited a comparable performance pattern, also demonstrated potential limitations in its ability to represent complex data structures. SVM is a dependable option for this application, as it boasts a performance level of over 84%.

The k-fold value of 10 implemented to C4.5 and DT models demonstrated superior performance in identifying underlying patterns in data, even with limited training data. However, using a larger k-fold value led to a small decrease in KNN performance, suggesting that surpassing a specific threshold for neighborhood size could impede its ability to identify local patterns. ANN showed slight improvements but still lagged behind top-performing algorithms, suggesting intrinsic constraints in the system's architecture or learning dynamics. NB persistent underperformance underscores the need to understand its underlying assumptions and potential limits when using it on datasets with feature dependencies. SVM demonstrated consistent performance, surpassing 84% across all criteria, solidifying its reputation as a reliable classifier for evaluating CPO quality.

This study has identified DT and C4.5 as the most effective machine learning algorithms for assessing the quality of CPOs. These algorithms consistently outperform other options, such as KNN and NB, due to their capacity to effectively represent complex data. SVM continues to be a viable alternative. The study results emphasize the importance of machine learning in improving quality evaluation in the palm oil industry. Future research should investigate ensemble methods, expand datasets, and understand the factors contributing to quality changes to improve predictive accuracy.

4. CONCLUSION

This paper investigates the application of machine learning in evaluating CPO quality. The findings indicate that C4.5 and DT algorithms are more effective in assessing CPO quality. These algorithms demonstrate exceptional performance in representing intricate relationships between datasets, essential for precisely predicting CPO quality. The key to the effectiveness of KNN is the careful selection of the optimal k-value, which highlights the need for precise parameter tuning. Notwithstanding their widespread use, the ANN and NB algorithms have limitations in effectively capturing nuanced data patterns when evaluating CPO quality. The SVM, however, is strong and dependable, capable of managing intricate data distributions and avoiding overfitting. This study has important implications for the palm oil industry, allowing stakeholders to shift towards more objective, precise, and efficient systems for evaluating the quality of CPO. Meanwhile, more investigation is needed to confirm that these algorithms are, in fact, the best fit, to clarify how they work, and to look into combining multiple models for even greater accuracy. This paper is vital in leading the palm oil sector toward a more data-driven form of quality control.

ACKNOWLEDGEMENTS

Thanks to PT. Telen Prima Sawit Muara Bengkal mill, Indonesia for their assistance in this research.

FUNDING INFORMATION

The Faculty of Engineering at Mulawarman University in Indonesia supported this project financially (Grant No. 8963/UN17.9/PT.00.03/2024).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Novianti Puspitasari	✓			✓			✓	✓	✓			✓		✓
Ummul Hairah		✓		✓	✓					✓			✓	
Vina Zahrotun Kamila						✓		✓		✓	✓		✓	
Hamdani Hamdani		✓			✓	✓				✓				
Anindita Septiarini				✓	✓			✓		✓				
Amin Padmo Azam Masa		✓	✓				✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [NP], upon reasonable request.

REFERENCES

[1] A. Solichin, U. Hasanah, and Jayanta, “Development of prediction system for crude palm oil (CPO) production with time series data mining approach,” in *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, Nov. 2020, pp. 147–152, doi: 10.1109/ICIMCIS51567.2020.9354321.

[2] BPS, *Statistik kelapa sawit Indonesia 2020*. Jakarta, Indonesia: Badan Pusat Statistik, 2021.

[3] M. Waleed, T. W. Um, T. Kamal, and S. M. Usman, “Classification of agriculture farm machinery using machine learning and internet of things,” *Symmetry*, vol. 13, no. 3, pp. 1–16, Mar. 2021, doi: 10.3390/sym13030403.

[4] K. Xu *et al.*, “A new machine learning approach in detecting the oil palm plantations using remote sensing data,” *Remote Sensing*, vol. 13, no. 2, pp. 1–17, Jan. 2021, doi: 10.3390/rs13020236.

[5] Hamdani, A. Septiarini, and D. M. Khairina, “Model assessment of land suitability decision making for oil palm plantation,” in *2016 2nd International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment*, Oct. 2017, pp. 109–113, doi: 10.1109/ICSITech.2016.7852617.

[6] H. Ji, X. Liu, L. Wang, L. Fan, and S. Liu, “Image recognition of Chinese herbal medicine using adaptive gamma correction based on convolutional neural network,” in *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, 2024, pp. 1428–1433, doi: 10.1109/DDCLS61622.2024.10606798.

[7] N. Puspitasari, A. Septiarini, U. Hairah, A. Tejawati, and H. Sulastri, “Betel leaf classification using color-texture features and machine learning approach,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, pp. 2939–2947, Oct. 2023, doi: 10.11591/eei.v12i5.5101.

[8] U. Hairah, A. Septiarini, N. Puspitasari, A. Tejawati, H. Hamdani, and S. E. Priyatna, “Classification of tea leaf disease using convolutional neural network approach,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 3287–3294, Jun. 2024, doi: 10.11591/ijece.v14i3.pp3287-3294.

[9] B. S. Reddy and S. Neeraja, “Plant leaf disease classification and damage detection system using deep learning models,” *Multimedia Tools and Applications*, vol. 81, pp. 24021–24040, 2022, doi: 10.1007/s11042-022-12147-0.

[10] A. Septiarini, H. Hamdani, E. A. Syaifudin, E. Setyaningsih, D. Nurcahyono, and N. A. Hadiwijaya, “Machine vision approach using multi features for detection of oil palm stem disease,” in *2023 1st International Conference on Advanced Engineering and Technologies, ICONNIC 2023 - Proceeding*, Oct. 2023, pp. 49–54, doi: 10.1109/ICONNIC59854.2023.10467726.

[11] J. Giménez-Gallego, J. Martinez-del-Rincon, J. D. González-Teruel, H. Navarro-Hellín, P. J. Navarro, and R. Torres-Sánchez, “On-tree fruit image segmentation comparing mask R-CNN and vision transformer models. Application in a novel algorithm for pixel-based fruit size estimation,” *Computers and Electronics in Agriculture*, vol. 222, Jul. 2024, doi: 10.1016/j.compag.2024.109077.

[12] M. Hussain, L. He, J. Schupp, D. Lyons, and P. Heinemann, “Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples,” *Computers and Electronics in Agriculture*, vol. 207, 2023, doi: 10.1016/j.compag.2023.107734.

[13] I. Nyalala *et al.*, “Tomato volume and mass estimation using computer vision and machine learning algorithms: cherry tomato model,” *Journal of Food Engineering*, vol. 263, pp. 288–298, Dec. 2019, doi: 10.1016/j.jfoodeng.2019.07.012.

[14] S. Jana, R. Parekh, and B. Sarkar, “A de novo approach for automatic volume and mass estimation of fruits and vegetables,” *Optik*, vol. 200, Jan. 2020, doi: 10.1016/j.ijleo.2019.163443.

[15] M. Mohd Ali, N. Hashim, and A. S. A. Hamid, “Combination of laser-light backscattering imaging and computer vision for rapid determination of oil palm fresh fruit bunches maturity,” *Computers and Electronics in Agriculture*, vol. 169, Feb. 2020, doi: 10.1016/j.compag.2020.105235.

[16] S. K. Behera, A. K. Rath, and P. K. Sethy, “Maturity status classification of papaya fruits based on machine learning and transfer learning approach,” *Information Processing in Agriculture*, vol. 8, no. 2, pp. 244–250, Jun. 2021, doi: 10.1016/j.inpa.2020.05.003.

[17] K. Tan, W. S. Lee, H. Gan, and S. Wang, "Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes," *Biosystems Engineering*, vol. 176, pp. 59–72, Dec. 2018, doi: 10.1016/j.biosystemseng.2018.08.011.

[18] Y. Kittichotsatsawat, N. Tippayawong, and K. Y. Tippayawong, "Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18635-5.

[19] H. Azgomi, F. R. Haredasht, and M. R. S. Motlagh, "Diagnosis of some apple fruit diseases by using image processing and artificial neural network," *Food Control*, vol. 145, 2023, doi: 10.1016/j.foodcont.2022.109484.

[20] S. Prabu, B. R. T. Bapu, S. Sridhar, and V. Nagaraju, "Tea plant leaf disease identification using hybrid filter and support vector machine classifier technique," *Intelligent Systems Reference Library*, vol. 215, pp. 117–128, 2022, doi: 10.1007/978-3-030-90119-6_10.

[21] S. Adige, R. Kurban, A. Durmuş, and E. Karaköse, "Classification of apple images using support vector machines and deep residual networks," *Neural Computing and Applications*, vol. 35, no. 16, pp. 12073–12087, 2023, doi: 10.1007/s00521-023-08340-3.

[22] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 23901–23928, 2024, doi: 10.1007/s11042-023-16194-z.

[23] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, 2021, doi: 10.1016/j.gsf.2020.03.007.

[24] F. Kitzler, H. Wagentristl, R. W. Neugschwandtner, A. Gronauer, and V. Motsch, "Influence of selected modeling parameters on plant segmentation quality using decision tree classifiers," *Agriculture*, vol. 12, no. 9, 2022, doi: 10.3390/agriculture12091408.

[25] D. L. Bersabal, J. L. Usa, E. R. Arboleda, and E. M. Galas, "Coffee bean recognition using shape features using decision trees and ensemble classifiers," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 4921–4924, 2020.

[26] Anjna, M. Sood, and P. K. Singh, "Hybrid system for detection and classification of plant disease using qualitative texture features analysis," *Procedia Computer Science*, vol. 167, pp. 1056–1065, 2020, doi: 10.1016/j.procs.2020.03.404.

[27] G. Saleem, M. Akhtar, N. Ahmed, and W. S. Qureshi, "Automated analysis of visual leaf shape features for plant classification," *Computers and Electronics in Agriculture*, vol. 157, pp. 270–280, 2019, doi: 10.1016/j.compag.2018.12.038.

[28] S. Wijaya and F. Fauziah, "Analysis of the comparison between linear regression, random forest, and logistic regression methods in predicting crude palm oil (CPO) price," *Brilliance: Research of Artificial Intelligence*, vol. 3, no. 2, pp. 343–350, Dec. 2023, doi: 10.47709/brilliance.v3i2.3334.

[29] K. Aqbar and R. A. Supomo, "Performance analysis of LSTM and XGBoost models optimization in forecasting crude palm oil (CPO) production at palm oil mill (POM)," *International Journal of Computer Applications*, vol. 185, no. 17, pp. 37–44, Jun. 2023, doi: 10.5120/ijca2023922890.

[30] N. A. Latif *et al.*, "Predicting heritability of oil palm breeding using phenotypic traits and machine learning," *Sustainability*, vol. 13, no. 22, Nov. 2021, doi: 10.3390/su132212613.

[31] M. F. Faruque, Asaduzzaman, and I. H. Sarker, "Performance analysis of machine learning techniques to predict diabetes mellitus," in *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, Feb. 2019, pp. 1–4, doi: 10.1109/ECACE.2019.8679365.

BIOGRAPHIES OF AUTHORS

Novianti Puspitasari received the B.Sc. degree in Informatics Engineering from the Universitas Islam Indonesia, and the M.Eng. degree in Information Technology from the Gadjah Mada University, Indonesia. She is currently a lecturer at the Department of Informatics, Mulawarman University. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), Indonesian Computer, Electronics, Instrumentation Support Society (IndoCEISS), Association of Computing and Informatics Institutions Indonesia (APTIKOM) societies and The Institution of Engineers Indonesia (PII). She has authored or coauthored more than 70 publications with 5 H-index. Her research interest is in data science and analytics, artificial intelligence, and machine learning areas. She can be contacted at email: novipuspitasari@unmul.ac.id.

Ummul Hairah is member of Institute of Electrical and Electronics Engineers (IEEE), and member of Association of Computing and Informatics Institutions Indonesia (APTIKOM) societies. Currently, she is actively teaching and researching at the Department of Informatics, Mulawarman University. As a writer on several journals and conferences with more than 40 publications, she focuses her research on database, information system and artificial intelligence. She can be contacted at email: ummul.hairah@fkti.unmul.ac.id.

Vina Zahrotun Kamila is graduated with a bachelor's degree in Informatics and a master's degree in Information Systems. She holds certificates related to governance and auditing as well as several other professional certificates. She is now working as lecturer and researcher in Department of Information System of Mulawarman University, Samarinda, East Kalimantan, Indonesia. Her research areas of interest are related to information system and evaluation of its use especially in education and small industry. She can be contacted at email: vinakamila@ft.unmul.ac.id.

Hamdani Hamdani is a professor in Department of Informatics at Mulawarman University, Indonesia. He is a lecturer and researcher since 2005 at Mulawarman University, Indonesia. His research interests lie in the field of artificial intelligence, especially pattern recognition, decision support system, and expert system. He received her bachelor's degree in 2002 from Ahmad Dahlan University, Indonesia, his master's degree in 2009 from Gadjah Mada University, Indonesia, and her doctoral degree in computer science in 2018 from Gadjah Mada University Indonesia. He can be contacted at email: hamdani@unmul.ac.id.

Anindita Septiarini is a Professor at the Department of Informatics at Mulawarman University, Indonesia. She holds a Doctoral degree in Computer Science from Gadjah Mada University, Indonesia, specializing in image analysis. She is also a researcher and got a grant from the Ministry of Education, Culture, Research, and Technology of Indonesia from 2016 until the present. Her research interests lie in artificial intelligence, especially pattern recognition, image processing, and computer vision. She has received national awards such as scientific article incentives from the Ministry of Education, Culture, Research, and Technology of Indonesia in 2017 and 2019. She held several administrative posts with the Department of Informatics, Mulawarman University, Indonesia, from 2018 to 2020, including the head of department and the head of laboratory. She can be contacted at email: anindita@unmul.ac.id.

Amin Padmo Azam Masa received the M.Cs. degree in Computer Science from the Gadjah Mada University, Indonesia. He is currently a lecturer at the Information System, Mulawarman University. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), Indonesian Computer, Electronics, Instrumentation Support Society (IndoCEISS), and Association of Computing and Informatics Institutions Indonesia (APTIKOM) societies. His research interest is in image processing, artificial intelligence, big data, and machine learning areas. He can be contacted at email: aminpadmo@unmul.ac.id.