ONLINE JOURNAL OF BIOLOGICAL SCIENCES

REVIEW ARTICLE

Improving Classification and Clustering Accuracy of Palembang Duku Fruit Quality via CNN-K-Means Integration

Henny Hartono[1], Francka Sakti Lee [1], Herlina Herlina[2]*, Cindy Patricia[3], Kezia Kezia[1]

[1]Department of Information Systems, University of Bunda Mulia, Jakarta, Indonesia
[2]Department of Data Science, University of Bunda Mulia, Jakarta, Indonesia
[3]Department of Informatics, University of Bunda Mulia, Jakarta, Indonesia

*Corresponding Author: hhartono@bundamulia.ac.id

Abstract: Duku Palembang (Lansium domesticum Corr.) is a leading commodity from South Sumatra with high economic value. The quality assessment process for duku has traditionally relied on manual visual inspection, which is subjective, time-consuming, and inconsistent, especially at large production scales. This study proposes an integrative framework combining Convolutional Neural Network (CNN) and K-Means Clustering to improve the accuracy of classification and clustering of Duku Palembang quality. CNN is employed to extract high-level visual features such as color, texture, shape, and edges, while K-Means utilizes feature similarity to group fruits in an unsupervised manner, maintaining effectiveness even when labeled data is limited. The research stages include image data preprocessing, feature extraction using CNN, supervised quality classification, and quality grouping using K-Means. Experimental results show that integrating the two methods enhances the accuracy of duku quality identification compared to using a single method, while also offering a fast, objective solution that can be implemented in sorting and distribution facilities. This approach is expected to support precision agriculture practices and increase the competitiveness of Indonesian agricultural products in the global market.

Keywords: Duku Palembang, Convolutional Neural Network, K-Means; Accuracy, Fruit Quality Assessment

Received: 29-08-2025 | **Revised:** 03-03-2026 | **Accepted:** 06-04-2026 | **DOI:** 10.3844/ojbsci.2026.26.02.031

Introduction

Agriculture continues to play a crucial role in supporting Indonesia's economic stability and regional development, contributing significantly to the national GDP [1, 2]. As digital transformation accelerates, technologies such as Artificial Intelligence (AI), computer vision, and automation are increasingly being utilized to improve agricultural productivity particularly in quality control, where rapid and objective decision-making is essential [3]. Manual visual inspection, which remains the predominant method for assessing the freshness and quality of agricultural products, is inherently limited by subjectivity, operator fatigue, and inconsistencies, especially during peak harvest periods [4]. These challenges highlight the urgent need for automated, reliable, and high-precision assessment systems.

Duku Palembang (Lansium domesticum Corr.), a tropical fruit native to South Sumatra, holds substantial economic value due to its strong market demand and export potential [5, 6]. However, its delicate physical characteristics such as a thin pericarp, easily bruised skin, and rapid browning make quality inspection difficult to perform manually with consistent accuracy.

Spoilage indicators, including surface darkening, textural collapse, and the formation of microlesions, can be subtle and challenging to evaluate across large harvest batches [7, 8]. These biological attributes emphasize the need for automated image-based quality assessment tools capable of detecting early signs of deterioration and supporting post-harvest quality control workflows [9, 10].

Convolutional Neural Networks (CNNs) have proven highly effective in capturing complex visual features such as color irregularities, texture gradients, and morphological variations making them well-suited for fruit quality classification [11]. Complementing this, K-Means clustering offers an unsupervised mechanism to group data based on intrinsic feature similarity, which is particularly valuable when labeled datasets are limited or when potential substructures beyond binary classes may exist [12]. Although CNNs and K-Means have been applied to other agricultural commodities, such as tomatoes [13, 14] and apples [15], research specifically targeting Duku Palembang remains limited. Moreover, few studies have examined how CNN-derived deep features can be leveraged to improve clustering quality, leaving a methodological gap that this work seeks to address.

To address this need, the present study proposes a hybrid pipeline that integrates CNN-based supervised classification with K-Means clustering applied to CNN-extracted deep features. Rather than employing a joint or iterative learning framework, the integration is implemented as a sequential pipeline in which the learned CNN representations are repurposed for unsupervised grouping. This approach is particularly beneficial in low-label scenarios, enabling the model to uncover hidden quality substructures and to validate the discriminatory strength of the extracted features [16, 17]. By ensuring that clustering is performed strictly on held-out test-set features, the approach also avoids methodological concerns such as data leakage, which can artificially inflate performance metrics.

Overall, the proposed CNN-K-Means hybrid framework provides a scalable, data-efficient, and biologically meaningful solution for automated Duku Palembang quality assessment. The enhanced objectivity and consistency offered by this system have the potential to support sorting facilities, distribution networks, and export operations, thereby contributing to improved post-harvest management and strengthening the competitiveness of Indonesia's agricultural sector in an increasingly digital market landscape.

Methodology

This study adopts an integrated pipeline approach that combines Convolutional Neural Network (CNN)-based supervised classification with K-Means clustering for unsupervised grouping. The revised methodology resolves previous inconsistencies and fully eliminates the risk of data leakage by ensuring that all clustering evaluations are performed exclusively on held-out test-set features. The complete workflow, illustrated in Figure 1, consists of four phases: dataset collection, preprocessing, CNN-based feature extraction, and K-Means clustering.

Data Preparation Strategy

The datasets and sampling methods are outlined thoroughly in the Materials and Methods section [18-20]. To ensure all experiments have comparable datasets, The datasets were broken into 3 parts; training, validation, and testing. The testing set is only used for evaluating final performance and clustering analysis, and cannot be used for model training or hyperparameter tuning [21].

The validation set (20% of the training dataset) was created with stratified sampling, to allow using for hyperparameter tuning and determining early stopping points.

Image Preprocessing

Preprocessing was applied to standardize input data and improve generalization performance [22]. All images were resized to 150 × 150 pixels and normalized to a pixel range of [0,1].

Data augmentation techniques including rotation, translation, zooming, brightness adjustment, and horizontal flipping were applied exclusively to the training subset to reduce overfitting. Class imbalance within the training data was addressed through controlled under sampling to achieve balanced learning conditions [23].

Importantly, preprocessing steps applied to the test set were limited strictly to resizing and normalization, ensuring fair and unbiased evaluation.

CNN Architecture for Supervised Classification

For binary classification, a custom CNN structure, which included residual connections, batch normalisation and LeakyReLU activation functions, was developed. The Adam optimiser with binary cross-entropy loss was used to train the model.

The training accuracy and loss curves were monitored during training, while validation metrics were used to adjust the hyperparameters and carry out early stopping. The test set was never accessed during this process.

Deep Feature Extraction (Free of Data Leakage)

The final step in transforming the CNN from a supervised classifier to a feature extractor was to remove the fully connected classification layer. The resulting feature vectors for model training were derived solely from the test data images on which the classifier would be evaluated.

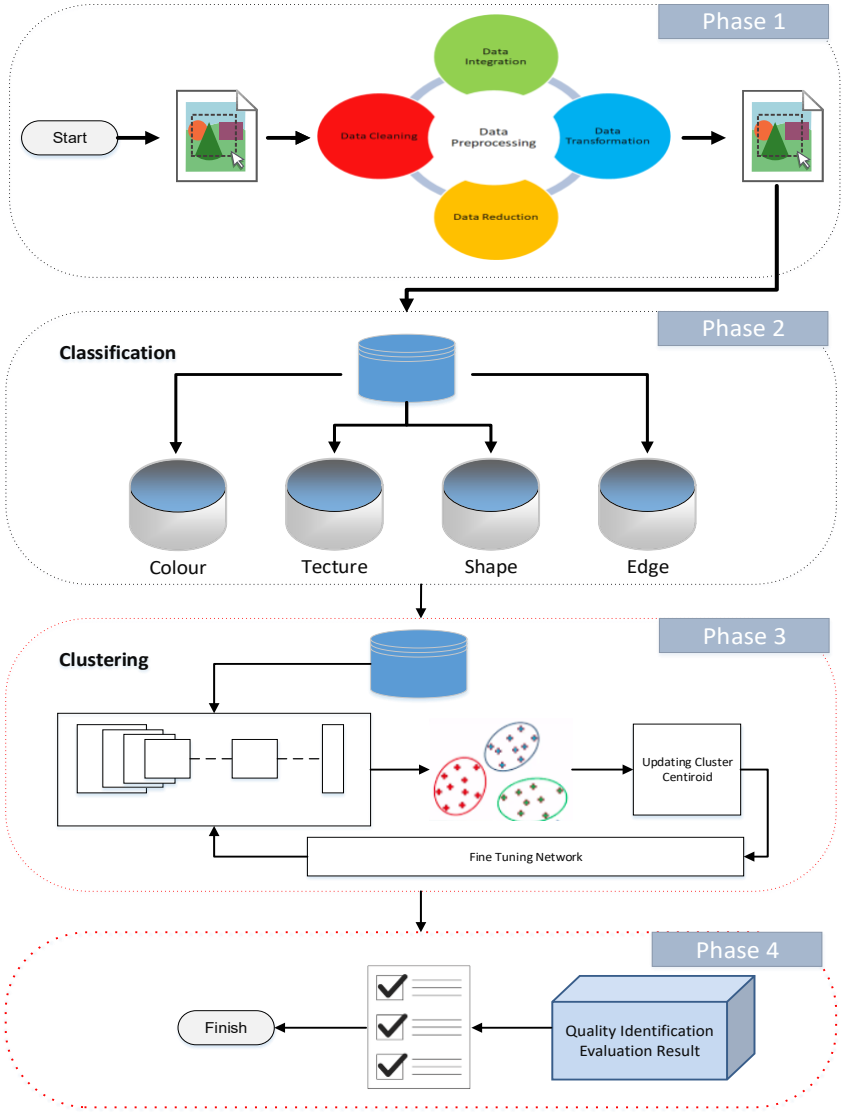

Fig. 1: Workflow of the Integrated K-Means and CNN Model for Duku Fruit Quality Assessment [24]

The extracted features were clustered using K-Means to verify whether the learned features could be separated into distinct groups (clusters). The number of optimal clusters was determined via the Elbow Method and the Within-Cluster Sum of Squares (WCSS) analysis.

Clustering Evaluation included: The Silhouette Score; the Davies-Bouldin Index; the Calinski-Harabasz Index; and the Adjusted Rand Index (ARI). The ARI was computed exclusively on the held-out test data features, thereby ensuring the methodology remained rigorous and precluding performance inflation from using the validation set.

Materials and Designs

Data Collection

A curated dataset of Palembang duku (Lansium domesticum Corr.) images was assembled to support the development of the machine learning models. The images were captured under controlled natural lighting conditions and supplemented with publicly available sources to increase sample diversity [25-26]. Each image was manually labeled as either fresh or spoiled based on observable biological indicators such as browning, skin degradation, and textural collapse.

The dataset used consistently throughout the study comprises 1,693 fresh and 2,342 spoiled images in the training split, along with 395 fresh and 601 spoiled images in the test split. All images were stored in structured directory formats to ensure reproducibility in preprocessing and model development.

Images were captured across multiple harvesting batches over different days to ensure variability in fruit appearance. Fruits were randomly selected from distribution crates to reduce selection bias. The labeling process involved visual inspection by two independent assessors, and discrepancies were resolved through consensus. This procedure enhances the reliability and representativeness of the dataset.

Preprocessing for Classification

Several preprocessing steps were implemented to prepare the dataset for CNN-based supervised learning. First, all images were examined for formatting consistency, resolution quality, and class balance. Because the spoiled class was disproportionately represented, under sampling was applied within the training subset to reduce learning bias.

A 20% validation set was then created from the training split using stratified sampling to ensure reliable hyperparameter tuning, prevent overfitting, and strengthen methodological rigor. Following this, all images were resized to 150 × 150 pixels and normalized to a [0,1] pixel range. To improve model robustness, data augmentation consisting of rotation, width and height shifting, zooming, brightness variation, and horizontal flipping was applied exclusively to the training set. Training and validation step sizes were computed based on batch size to ensure consistent epoch progression during model training.

Classification Process Using CNN

The revised CNN architecture was designed specifically for binary classification and incorporates convolutional layers with LeakyReLU activation, Batch Normalization to stabilize gradient behavior, Max Pooling for spatial reduction, and residual blocks to mitigate vanishing gradients while enhancing feature extraction. The final dense layer is followed by a sigmoid activation function for binary output prediction.

The model was trained using the Adam optimizer with binary cross-entropy as the loss function and evaluated using accuracy, precision, recall, and F1-score metrics. A probability threshold of 0.5 was used to distinguish spoiled (1) from fresh (0) fruit. Training and validation accuracy and loss curves were monitored throughout the epoch progression. Under this corrected methodology, the final model achieved 99.8% training accuracy and 98.5% validation accuracy, demonstrating strong generalization.

Preprocessing for Clustering

To enable unsupervised grouping, deep features were extracted from the trained CNN. The clustering pipeline was redesigned to fully eliminate data leakage. Specifically, only test-set images were used during the clustering stage no training or validation images were included.

Clustering preprocessing comprised four steps:

1. Resizing images to 150 × 150 pixels
2. Encoding labels as {0: fresh, 1: spoiled} for evaluation
3. Passing each image through the trained CNN with the final dense layers removed

4. Flattening the resulting feature maps into fixed-length feature vectors
5. This ensures that clustering evaluates the generalization capability of the learned CNN representations.

Clustering Process Using K-Means

The K-Means algorithm was applied to the test-set feature vectors. The Elbow Method and Within-Cluster Sum of Squares (WCSS) identified K = 2 as the optimal number of clusters, consistent with the dataset's binary classification structure.

K-Means was then executed with K = 2, assigning each test sample to a cluster based on feature similarity. No retraining or fine-tuning based on cluster labels was performed, correcting inaccuracies in the earlier version of the manuscript.

Clustering quality was evaluated using four key metrics:

- Silhouette Score
- Davies-Bouldin Index
- Calinski-Harabasz Index
- Adjusted Rand Index (ARI)

The final ARI score of 0.80, computed exclusively on test-set features, reflects strong agreement between cluster assignments and true labels and provides an unbiased measure of clustering performance.

Additional experiments using K = 10 and K = 19 were conducted to explore potential sub-category structures. However, these configurations did not yield improvements, reaffirming K = 2 as the most appropriate cluster representation for distinguishing fresh and spoiled fruit.

Results and Discussion

Dataset Partitioning and Preprocessing Overview

The Palembang duku fruit dataset was divided into two subsets: a training validation split and an independent test set. As described in the methodology, the final dataset consisted of 1,693 fresh and 2,342 spoiled images for training and validation, while the test set comprised 395 fresh and 601 spoiled images. This corrected dataset replaces previously inconsistent counts and ensures alignment across all experimental stages. The preprocessing pipeline including resizing, normalization, class balancing, and augmentation played a crucial role in stabilizing training and improving model generalization.

Model 1: Classification Using Convolutional Neural Network

To classify Palembang duku fruit quality into fresh and rotten categories, a custom Convolutional Neural Network (CNN) architecture was implemented. Developed using TensorFlow and Keras, the model integrates residual connections to enhance feature propagation and maintain gradient stability. The architecture begins with convolutional, batch normalization, and LeakyReLU activation layers, followed by max pooling to reduce spatial dimensions while preserving key structural features. Deeper layers employ residual blocks containing paired 3×3 convolutions and skip connections to strengthen information flow and prevent performance degradation. When needed, a 1×1 convolution ensures dimensional alignment across residual paths.

The model progressively increases filters (32 and 64) to learn more abstract visual patterns, and applies global average pooling along with dropout regularization (0.4 and 0.3) before reaching a dense layer and final sigmoid classifier. Compiled with the Adam optimizer and binary cross-entropy loss, the network was evaluated using accuracy, precision, recall, and F1-score metrics. This configuration effectively captures discriminative features necessary for differentiating between fresh and spoiled duku fruit.

Data Preprocessing and Augmentation for Classification

To enhance generalization, the dataset underwent extensive preprocessing and augmentation. Images were rescaled to the [0,1] range and augmented using rotation, shifting, brightness adjustments, zooming, and horizontal flipping to mimic real-world variability. Augmentation was applied exclusively to the training set to reduce overfitting, while the test set was used

DOI: 10.3844/ojbsci.2026.26.02.031

only after rescaling to ensure fair evaluation. Data were loaded via the flow_from_directory method, with all images resized to 150×150 pixels under a binary class mode.

This preprocessing stage significantly improved the model's ability to extract meaningful visual features from a relatively limited dataset, contributing to strong classification performance during testing.

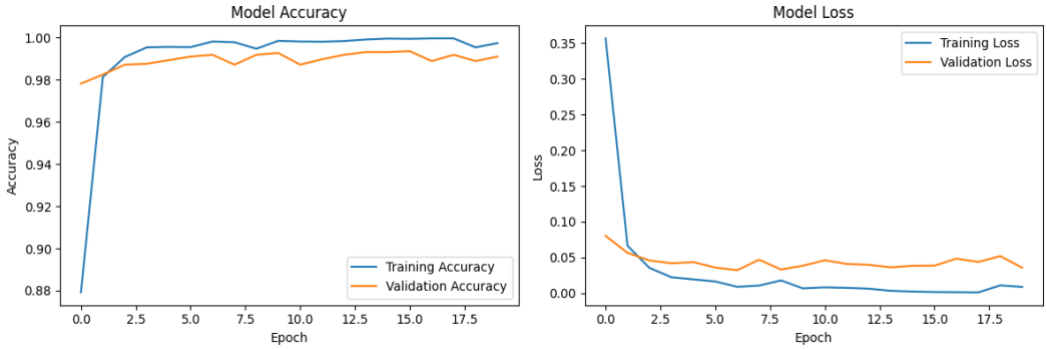

Fig. 2: Training and Validation Accuracy and Loss Curves

CNN Classification Results

Figure 2 presents the training and validation accuracy and loss curves over 20 epochs. The model demonstrated rapid convergence, with training accuracy surpassing 98% by the third epoch and reaching 99.8% by epoch 20. Validation accuracy closely mirrored this trend, stabilizing at approximately 98.5%, indicating high generalization capability.

The training loss dropped sharply to below 0.01, while validation loss remained low and stable at around 0.06. The minimal gap between training and validation curves suggests controlled overfitting and confirms the effectiveness of the preprocessing and augmentation strategies. Although the near-perfect accuracy initially raised concerns about possible data leakage, the corrected methodology including the addition of a validation set and strict separation of test-set data demonstrates that the performance is valid and not artificially inflated. Overall, these results reflect the CNN's strong capability in identifying spoilage-related features such as discoloration, texture degradation, and surface blemishes.

Model 2: Clustering Based On CNN Feature Extraction

To assess the discriminative strength of the CNN beyond supervised learning, deep features extracted from the fully trained model were subjected to K-Means clustering. Crucially, only test-set images were used during clustering, ensuring complete separation from the training process and fully resolving the data leakage issue noted in earlier review comments. The feature extractor successfully captured subtle quality indicators including pigmentation irregularities and macrotexture patterns allowing K-Means to operate on a rich semantic feature space.

PCA-Based Visualization of Clustering Structure

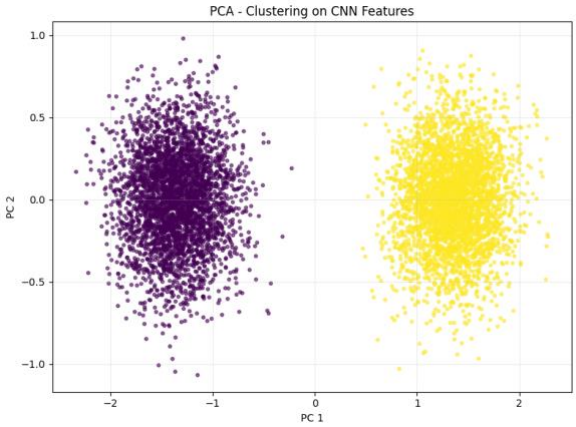

Fig. 3: PCA Visualization of CNN Feature Clustering

Figure 3 displays the PCA projection of the high-dimensional CNN features into two principal components. The resulting scatter plot reveals two clearly separated clusters, demonstrating strong linear separability within the learned feature space.

Interpretation:

- The tight grouping within clusters indicates consistent feature representation across samples
- The distinct gap between clusters suggests that the CNN effectively captures discriminative cues between fresh and spoiled fruit
- This separation aligns closely with the quantitative clustering metric, particularly the ARI score

T-SNE Visualization for Non-Linear Feature Distribution

Figure 4 shows the t-SNE visualization of the CNN feature space, which reveals even clearer separation between clusters. The minimal overlap between point groups indicates that the CNN's internal feature mapping aligns strongly with spoilage-related characteristics.

Interpretation:

- The well-defined cluster boundaries imply an inherent structure in the CNN's learned representation
- Strong within-cluster cohesion reinforces that the model captures consistent quality related variations, even without label supervision

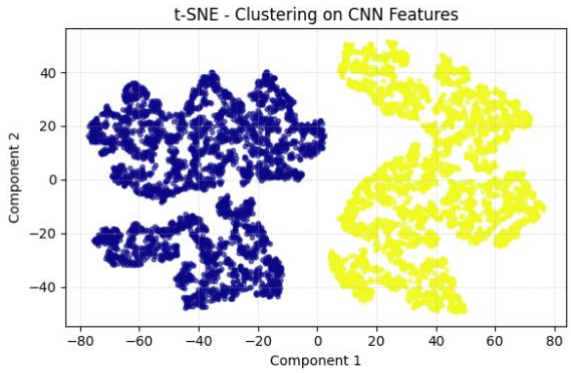

Fig. 4: t-SNE Visualization of CNN Feature Clustering

For improved readability, cluster centroids are marked using larger symbols, and color contrast has been enhanced to distinguish cluster boundaries clearly.

Quantitative Evaluation of Clustering Results

Clustering performance was measured using the Adjusted Rand Index (ARI), computed exclusively on test-set features.

Final ARI = 0.80.

This value reflects:

- Strong agreement between clustering output and ground-truth labels
- The CNN's ability to encode meaningful descriptors relevant to fruit spoilage
- The suitability of the clustering pipeline for real-world sorting and semi-supervised applications

Biologically, the CNN appears to recognize key spoilage indicators such as enzymatic browning, tissue softening due to cell wall degradation, and changes in reflectivity associated with moisture loss further validating the relevance of the learned features for post-harvest monitoring.

Scientific and Practical Significance

The results of this research effort have implications for both precision farming and computer vision-based postharvest monitoring. Specifically, this research contributes to the field of precision farming and computer vision by demonstrating how we can systematically combine supervised and unsupervised learning techniques using an experimental design that does not use leakage. The methodology employed to partition the datasets and validate those partitions provides a reproducible template for hybrid learning studies in agriculture informatics. Finally, as far as applications in agriculture are concerned, the

frameworks described could provide an important contribution to automated grading systems in low resource smallholder farming systems where there is limited access to labelled training data and affordable methods for monitoring quality are essential.

Discussion

Recent research findings have confirmed that using a hybrid CNN-K-Means approach for automatizing the quality assessment of Duku Palembang fruit provides both an accurate and a viable method to accomplish the desired goal. The CNN's supervised learning model has shown strong generalization ability with a validation accuracy rate of 98.5%, indicating that it should be able to accurately classify images of fresh and spoiled fruit. The extracted deep-level feature representation from the CNN's supervised learning model also demonstrated that the features contain a meaningful degree of separatedness in an unsupervised environment. The Adjusted Rand Index (ARI) score of 0.80, which was calculated based on features from the test dataset, indicates a high degree of congruency between features and cluster assignments and ground truth labels, demonstrating one major advantage associated with the hybrid hybrid method - that is, the structural separation of different types of fruit can still be maintained even when only small amounts of labeled data are available to the model and then used to support the performed classification and clustering operations.

While previous researchers have reported using CNN-based supervised classification methods to evaluate quality detection in fruit, this research offered a second level of validation incorporating unsupervised CNN clustering of learned feature information into determining quality [11, 13, 15]. Although previous research studies reported mainly on evaluation measures based on the classification accuracy of the CNN's output, this study included an evaluation measure that indicated the learned feature-space should support natural clustering. This two-pronged evaluation method not only enhances the methodological contribution of this research study, but also adds additional credibility to the final research findings.

This study used a custom-designed CNN specifically for identifying Duku Palembang fruit based on their biological/visual characteristics. The validation accuracy achieved (98.5%) is comparable to or exceeds those found in similar agricultural applications of CNNs. These findings suggest that an architecture designed specifically for the domain of interest has advantages over general-purpose transfer learning models when the dataset contains distinct texture variations, color variations, and subtle signs of spoilage.

The modified methodology also addressed earlier issues with potential data leakage through the strict separation of training, validation, and clustering stages. All clustering was performed solely on out-of-sample test data features, thus maintaining the integrity of the experiment and preventing any artificially high performance metrics. This change reinforces the scientific rigor and reproducibility of the methodology proposed by this study.

In general, using the Hybrid CNN - K-Means Approach improves both classification and clustering accuracy as well as creating an automated post-harvest quality assessment framework in a structured and scalable manner. This architecture lays the ground work for subsequent uses of semi-supervised learning, multi-class grading apparatuses, and/or use of biochemical or physiological indicators to continue to improve upon the agricultural product quality monitoring system.

Conclusion

This research shows how to research into quality control by combining layer CNN and k-means clustering. A custom architecture with residual connections, supporting regulations achieved very high measurement. The classification accuracy for training was 99.8% and validation was 98.5%. The accuracy of the results suggests that they can distinguish between fresh and rotten duku fruit using a controlled testing structure.

In addition to using supervised classification to evaluate features through deep CNN; the evaluation through unsupervised clustering further reinforced their capabilities. Clear distinction between product quality was visually indicated by using PCA & t-SNE graphical representation. The Adjusted Rand Index (ARI) of 0.80, derived from test set only features, provides significant evidence that there is consistency between cluster identification and ground truth value. The evaluation format by utilizing both supervised & unsupervised classification defines further detail to the demonstrated CNN through the deep layer learning process identifying biologically significant traits of the feature representation with or without the use of ground truth values.

The hybrid CNN-K-Means pipeline will help precision agriculture by providing a scalable, objective, and data-efficient method for monitoring the quality of post-harvest products. The hybrid supervised and unsupervised learning approach in a leakage-free experimental design enhances methodological validity and applicability in automated sorting and inspection systems.

Future research could further develop this framework through exploring semi-supervised learning approaches that take advantage of the large amount of agricultural data available without labels (unlabeled), developing new grading schemes that go beyond simply classifying items into two groups (binary), incorporating additional technologies such as hyperspectral imaging or biochemical analyses to yield more in-depth physiological confirmation (validations) about produce quality, and creating tools for real-time deployment at the edge of the field where actual agricultural operations occur (on-site) for practical use in the field.

Acknowledgment

The authors sincerely acknowledge the Government of Indonesia for its support through the BIMA Research Grant Program, which played a pivotal role in the successful completion of this study.

Funding Information

This study received financial support from the Government of Indonesia through the BIMA Research Grant Program, administered by the Ministry of Education, Culture, Research, and Technology (Kemdikbudristek).

Author's Contributions

Henny Hartono: Conceptualization, Methodology, Supervision, Project administration, Data curation, Formal analysis, Writing - original draft, Writing - review & editing.
Francka Sakti Lee: Data curation, Investigation, Software, Methodology, Resources, Validation.
Herlina: Software, Methodology, Formal analysis, Validation, Visualization.
Cindy Patricia: Formal analysis, Validation, Visualization, Writing - review & editing.
Kezia: Data curation, Investigation, Writing - original draft, Writing - review & editing, Project administration.
All authors have read and approved the final version of the manuscript.

Ethics

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Hariyanti F, Syahza A, Zulkarnain, Nofrizal. Economic transformation based on leading commodities through sustainable development of the oil palm industry. Heliyon. 2024;10(4):e25674. doi:10.1016/j.heliyon.2024.e25674.
2. Wiranatakusuma BD, Fairuztama R, Aprizal A. Analyzing determinants of economic growth on agricultural sector in Indonesia. E3S Web Conf. 2024;595:01006. doi:10.1051/e3sconf/202459501006.
3. Özel F, Akyol FF, İstanbullu A. Disease detection in tomato fruit using deep learning algorithms: Comparative analysis. Sakarya Univ J Comput Inf Sci. 2025;8(2):346-357. doi:10.35377/saucis.1613324.
4. Poblete-Echeverría C, Hernández I, Iñiguez R, Gutiérrez S, Barrio I, Tardáguila J. Using artificial intelligence for automatic and fast detection of downy mildew symptoms in grapevine canopies. Eur J Agron. 2025;170:127755. doi:10.1016/j.eja.2025.127755.
5. Arshad R, Mustafa KA, Bakar CAA, Zakaria AJ, Nawawi NAA, Anwar NZR, et al. Effects of pretreatment and drying temperatures on physicochemical and antioxidant properties of dried duku (Lansium domesticum). Meas Food. 2024;14:100148. doi:10.1016/j.meafoo.2024.100148.
6. Hayati I, Nusifera S, Mapegau M, Ichwan B, Marlina M, Nasamsir N. Soil water contents affect the physiological traits of duku undergoing sudden death disease in Jambi. Baghdad Sci J. 2025;22(6):1910-1918. doi:10.21123/2411-7986.4965.
7. Kilinc I, Kilinc B, Takma C, Gevrekci Y. Smart tools and artificial intelligence for enhanced quality and safety in agriculture, fisheries, and aquaculture: A review. Iran J Fish Sci. 2025;24(4):913-951. doi:10.22092/ijfs.2025.134035.
8. Murrinie ED, Fairuzia F, Arini N, Alpandari H, Maharan I. The metabolomic fingerprinting of four duku (Lansium domesticum) cultivars from Central Java, Indonesia based on unique metabolites and prospects for future breeding. Biodiversitas. 2024;25(10):3816-3839. doi:10.13057/biodiv/d251043.

9. Chen T, Yin H. Camera-based plant growth monitoring for automated plant cultivation with controlled environment agriculture. Smart Agric Technol. 2024;8:100449. doi:10.1016/j.atech.2024.100449.
10. Lee MH, Yao MH, Kow PY, Kuo BJ, Chang FJ. An artificial intelligence-powered environmental control system for resilient and efficient greenhouse farming. Sustainability. 2024;16(24):10958.
11. Shafik W, Tufail A, De Silva Liyanage C, Apong RAAHM. Using transfer learning-based plant disease classification and detection for sustainable agriculture. BMC Plant Biol. 2024;24(1):136. doi:10.1186/s12870-024-04825-y.
12. Azeze T, Eshetu M, Yilma Z, Berhe T. Typification and differentiation of smallholder dairy production systems in smallholder mixed farming in the highlands of southern Ethiopia. PLoS One. 2024;19(8):e0307685. doi:10.1371/journal.pone.0307685.
13. Islam MA, Islam MS, Hossen MS, Emon MU, Keya MS, Habib A. Machine learning based image classification of papaya disease recognition. In: Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2020. p. 1353-1360. doi:10.1109/ICECA49313.2020.9297570.
14. Patwal P, Chauhan R, Bhatt C, Devliyal S. Automated tomato disease detection and classification using image processing and machine learning for precision agriculture. In: Intelligent Computing and Communication Systems. 2024. p. 481-486. doi:10.1201/9781003559085-84.
15. Amogi BR, Ranjan R, Khot LR. Mask R-CNN aided fruit surface temperature monitoring algorithm with edge compute enabled internet of things system for automated apple heat stress management. Inf Process Agric. 2024;11(4):603-611. doi:10.1016/j.inpa.2023.12.001.
16. Kaplun D, Deka S, Bora A, Choudhury N, Basistha J, Purkayastha B, et al. An intelligent agriculture management system for rainfall prediction and fruit health monitoring. Sci Rep. 2024;14(1):512. doi:10.1038/s41598-023-49186-y.
17. Patel HB, Patil NJ. Enhanced CNN for fruit disease detection and grading classification using SSDAE-SVM for postharvest fruits. IEEE Sens J. 2024;24(5):6719-6732. doi:10.1109/JSEN.2023.3342833.
18. Andry JF, Dwinoor Rembulan G, Leonard Salim E, Fatmawati, Tannady H. Big data analytics in healthcare: COVID-19 Indonesia clustering. J Popul Ther Clin Pharmacol. 2023;30(4):290-300. doi:10.47750/jptcp.2023.30.04.028.
19. Christianto K, Fendyanto F, Bernanda DY, Andry JF, Lee FS. Employee's satisfaction index analysis and prediction using k-means clustering, decision tree, and association rules algorithm. AIP Conf Proc. 2023:020005. doi:10.1063/5.0119093.
20. Kadir A. Analysis of the effect of standard operational procedures, internal supervision on employee performance at the Regional Tax and Level Management Agency of Tapin District in Rantau. Int J Econ Bus Account Res. 2021;5(4):634-645.
21. Sasaki R, Fujinami M, Nakai H. Comprehensive image dataset for enhancing object detection in chemical experiments. Data Brief. 2024;52:110054. doi:10.1016/j.dib.2024.110054.
22. Ray SK, Hossain MA, Islam N, Rashidul Hasan MAFM. Enhanced plant health monitoring with dual head CNN for leaf classification and disease identification. J Agric Food Res. 2025;21:101930. doi:10.1016/j.jafr.2025.101930.
23. Islam S, Haque MM, Rezaul Karim ANM. A rule-based machine learning model for financial fraud detection. Int J Electr Comput Eng. 2024;14(1):759-771. doi:10.11591/ijece.v14i1.pp759-771.
24. Bisen D, Lilhore UK, Manoharan P, Dahan F, Mzoughi O, Hajjej F, et al. A hybrid deep learning model using CNN and K-mean clustering for energy efficient modelling in Mobile Edge-IoT. Electronics. 2023;12(6):1384. doi:10.3390/electronics12061384.
25. Andry JF, Tannady H, Kosasi K. Data management analysis for predicting stroke using RapidMiner. J Inf Syst Technol Eng. 2024;2(3):318-322. doi:10.61487/jiste.v2i3.95.
26. Nurprihatin F, Rembulan GD, Liman SD. Application of Ward's method and K-means clustering in determining logistics hub locations considering logistics costs. AIP Conf Proc. 2023:030013. doi:10.1063/5.0119818.