BIO Web of Conferences **234**, 02018 (2026) https://doi.org/10.1051/bioconf/202623402018
FiSAED 2025

Image-based classification of Golden Melon ripeness using convolutional neural networks and data augmentation

Endang Purnama Giri[1*], *Agus* Buono[1], *Karlisa* Priandana[1], and *Dwi* Guntoro[2]

[1]Computer Science Study Program, School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia
[2]Department of Agronomy, Faculty of Agriculture, IPB University, Bogor 16680, Indonesia

Abstract. Accurate fruit maturity assessment is critical for postharvest quality control and optimal harvest scheduling. This study investigates the performance of machine learning and deep learning approaches for classifying Golden Alisha melon maturity using image data. A self-developed dataset comprising 230 labelled images was organized into four maturity stages: 47, 53, 60, and 67 days after planting (DAP). Four classification models were evaluated: Principal Component Analysis combined with Support Vector Machines (PCA–SVM), Principal Component Analysis with Neural Networks (PCA–NN), a Convolutional Neural Network (CNN), and a CNN enhanced with data augmentation. The augmentation strategy included random rotations (≤20°), horizontal and vertical shifts (≤10%), zooming (≤20%), and horizontal flipping applied during training. Performance was assessed using precision, recall, F1-score, and overall accuracy. Experimental results demonstrate that the augmented CNN achieved the highest accuracy of 86%, outperforming both PCA-based models and the baseline CNN. Confusion matrix analysis reveals that the 60 DAP stage is the most difficult to classify, likely due to visual similarity with adjacent ripening stages. This study provides a curated image dataset and confirms the effectiveness of data-augmented deep learning for robust agricultural maturity classification.

1 Introduction

Melon is a high-value horticultural commodity with increasing demand in both domestic and international markets. In Indonesia, melon production reached 92,432 tons in 2017, covering a harvested area of 5,879 hectares [1]. Among the cultivated varieties, Golden Alisha melon has gained significant attention due to its market acceptance, including export demand to the Middle East. The wide distribution range and extended supply chain require accurate production and harvest management to ensure consistent quality upon arrival at end consumers.

* Corresponding author: endang_pg@apps.ipb.ac.id

One of the most critical challenges in melon production is the accurate determination of fruit ripeness. Melon is classified as a climacteric fruit, in which ripening processes continue after harvest, including changes in sugar content, texture, and skin appearance. As a result, harvest decisions cannot rely solely on general ripeness categories but must be determined more precisely based on Days After Planting (DAP). This is particularly important for export-oriented production, where melons are typically harvested before full maturity to achieve optimal ripeness during distribution. Precise ripeness identification based on DAP plays a key role in minimizing post-harvest losses and maintaining product uniformity across the supply chain.

In recent years, Artificial Intelligence (AI) based approaches have shown promising results in automating fruit ripeness classification. Previous studies have explored machine learning techniques for melon ripeness assessment, demonstrating the feasibility of intelligent classification systems. However, many existing works rely on limited datasets or generalized ripeness labels, which may not adequately represent ripeness progression based on planting age. This limitation highlights the need for high-quality datasets and models that explicitly incorporate DAP-based ripeness stages.

To address this gap, this study develops a melon ripeness classification model based on Days After Planting using a newly constructed image dataset of Golden Alisha melons. The dataset was collected at the Agrotechno Park (ATP) Cikarawang, IPB University Experimental Farm, under controlled lighting and acquisition conditions to ensure data consistency. While designed primarily for image-based analysis, the dataset is structured to support future multimodal extensions, including acoustic and other sensory data.

Previous studies have demonstrated the effectiveness of CNN in image classification tasks wrote on [2]. Comparative study between Support Vector Machine (SVM) and neural Network (NN) wrote on [3]. On the other hand successful story about feature reduction by PCA wrote on [4]. In this research CNN model is proposed and evaluated, with its performance compared against conventional machine learning methods SVM and NN using Principal Component Analysis (PCA) for feature reduction. The results establish a strong visual-based baseline and highlight the advantages of deep learning for DAP-oriented ripeness classification in agricultural applications.

2 Melon ripeness and dataset acquisition procedure

This section describes the materials and methods used in this study, including the biological basis of melon ripeness classification based on Days After Planting (DAP), the image dataset acquisition procedure, and the classification methods employed for performance evaluation.

2.1 Melon ripeness classification

Melons (Cucumis melo) are climacteric fruits that continue to undergo physiological ripening after harvest, including increases in sugar content, flesh softening, and changes in aroma and texture. Consequently, harvest timing is a critical determinant of final fruit quality. Harvesting at an immature stage may result in insufficient sweetness and poor flavour, whereas delayed harvesting can cause excessive softening and reduced storability. To address this issue, Days After Planting (DAP) is widely used as a quantitative indicator of physiological maturity. Previous studies have demonstrated that melon fruits harvested at different DAP exhibit significant variations in sugar accumulation, flesh firmness, and ethylene production, making DAP a reliable proxy for ripeness stage.

DAP also provides an objective and reproducible basis for ripeness classification, which is particularly important for the development of data-driven agricultural systems. As a measurable parameter directly related to fruit physiological age, DAP reduces the subjectivity associated with visual inspection and enables consistent labelling for machine learning and deep learning models. Moreover, because melon is a climacteric fruit, harvesting based on DAP allows producers to target a pre-climacteric or early climacteric stage, ensuring that optimal ripeness is achieved during postharvest handling and distribution. This approach contributes to improved harvest scheduling and quality consistency across production cycles. Despite these advantages, existing melon ripeness assessment methods still rely heavily on manual inspection or destructive measurements, which are time-consuming, labour-intensive, and impractical for large-scale applications. Recent studies have explored machine learning techniques for fruit maturity classification; however, many of them are limited by small datasets, insufficient feature representation, or the lack of systematic comparison between conventional machine learning and deep learning approaches. In addition, the potential of data augmentation to enhance model robustness in melon maturity classification remains underexplored.

To address these gaps, this study proposes an image-based maturity classification framework for Golden Alisha melons using machine learning and convolutional neural network (CNN) models. The main contributions of this work are threefold: (1) the construction of a labelled image dataset representing multiple maturity stages based on DAP, (2) a comprehensive comparison between PCA-based machine learning methods and CNN-based deep learning models, and (3) an evaluation of the impact of data augmentation on classification performance. The results provide empirical evidence that deep learning with data augmentation offers superior performance for automated melon maturity assessment, highlighting its potential for practical deployment in intelligent agricultural systems.

This study focuses exclusively on the Golden Alisha melon cultivar, which is characterized by a round to slightly oval shape, smooth rind without netting, and a color transition from light green at the immature stage to bright yellow at full ripeness. Harvest timing plays a decisive role in determining fruit quality for this variety. In practice, harvesting is commonly performed prior to full physiological maturity, typically 5-7 days before complete ripeness [5], to allow sufficient time for postharvest handling processes such as sorting and transportation. Figure 1 shows Melon harvesting process. Harvesting at an excessively early stage may result in fruit with suboptimal sweetness and incomplete size development, whereas delayed harvesting can lead to overripe fruit with reduced shelf life and increased susceptibility to mechanical damage during distribution.

For Golden Alisha melons, maturity assessment is generally based on Days After Planting (DAP), which serves as a practical and quantitative indicator of ripeness stage. The classification of ripeness levels according to DAP provides a consistent framework for determining harvest readiness and ensuring quality uniformity, as summarized in previous studies [6]:

- 45 DAP: The fruit is classified as unripe, characterized by green and firm skin. At this stage, the melon is not suitable for harvesting and is primarily used for early growth and developmental monitoring.
- 53 DAP: This stage represents a transition toward ripening. The skin begins to show a slight yellow coloration, and the fruit texture starts to soften, making it suitable for pre-harvest evaluation.
- 60 DAP: The fruit reaches a near-optimal ripeness stage and is generally considered the most appropriate harvest time for long-distance distribution, as it has achieved sufficient physiological maturity while maintaining adequate firmness for transportation.
- 67 DAP: The fruit is fully ripe, exhibiting bright yellow skin, a strong sweet aroma, and very soft flesh. This stage is best suited for immediate consumption or local markets but presents a high risk of damage during long-distance transportation.

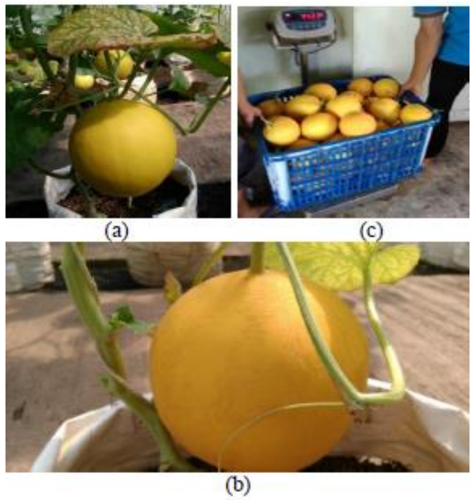

Fig. 1. Melon harvesting process (a) & (b) ripening and harvesting; (c) weighing harvested yield [5].

In the context of melon ripeness assessment, non-destructive techniques are considered highly advantageous, as they preserve fruit integrity and allow classification to be performed without the need for harvesting. These characteristics are well aligned with the requirements of modern agricultural systems, which emphasize efficiency, sustainability, and product quality preservation. Prayoga *et al.* [7] developed a non-destructive ripeness classification model for the Sky Rocket melon variety using a Support Vector Machine (SVM) approach. The proposed classification consisted of three ripeness categories, namely unripe (45–51 DAP), half-ripe (51–59 DAP), and ripe (61–70 DAP). The study utilized a dataset of 450 manually cropped melon images, from which grayscale histogram features were extracted using adaptive thresholding to represent skin texture density, a visual indicator associated with fruit maturity. The model achieved an overall classification accuracy of 76%.

Subsequent research by Saputra [8] demonstrated that the integration of Gray Level Co-occurrence Matrix (GLCM) feature extraction with SVM classification could further improve melon ripeness classification performance, achieving an accuracy of 80%, with precision and recall values of 81% and 80%, respectively. More recently, Alfarizy and Fitri [9] proposed a convolutional neural network (CNN)-based model for melon ripeness classification using image data categorized into unripe, ripe, and overripe classes. Their approach attained a high classification accuracy of 90%, highlighting the effectiveness of deep learning methods for automated ripeness detection in agricultural applications.

Despite these encouraging results, most existing studies have focused on Joey Rocket or other netted-skin melon varieties, which exhibit distinct visual and morphological characteristics compared to the Golden Alisha variety examined in this study. Consequently, classification models developed for netted melons cannot be directly transferred to smooth-skinned Golden Alisha melons without substantial adaptation, as the visual features used for classification are not fully representative. This limitation underscores the need for a dedicated investigation that specifically targets the Golden Alisha variety using an appropriate visual dataset.

A recent study involving the Golden melon variety was conducted by Arham [10], employing the YOLOv8 deep learning framework. The model achieved an accuracy of 91.2%, with precision and recall values of 0.89 and 0.90, respectively. Although YOLOv8, which is fundamentally based on convolutional neural network architectures, demonstrated strong performance, the ripeness classification was restricted to only two categories: ripe and unripe. Therefore, to enable finer-grained ripeness classification, such as classification based on Days After Planting (DAP), a more comprehensive modelling approach and deeper analysis are required.

2.2 Dataset acquisition procedures

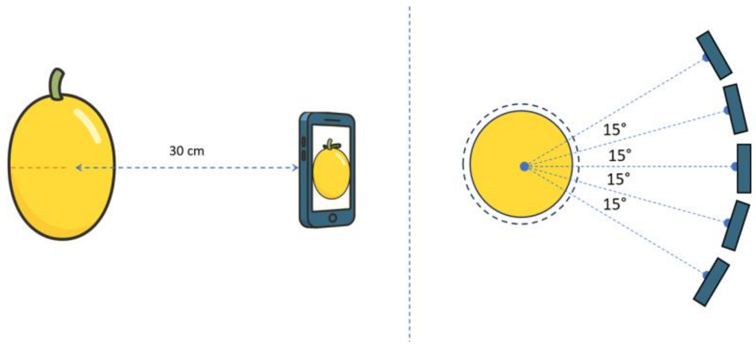

Fig. 2. Configuration for capturing image dataset

Figure 2 illustrates the data acquisition procedure for capturing images of Golden melon fruit while still attached to the plant using a smartphone camera. The acquisition protocol was defined as follows.

1. Fixed camera-to-fruit distance (30 cm). The smartphone was positioned at a constant distance of 30 cm from the melon surface, measured using a ruler or measuring tape. Although the fruit was not physically fixed, the operator ensured that this distance was consistently maintained for all image captures to preserve uniform image scale.
2. Camera alignment with the fruit equator. The camera lens was aligned with the equatorial plane (horizontal midpoint) of the melon. This alignment ensured consistent vertical positioning and minimized perspective distortion caused by capturing images from excessively high or low angles.
3. Five image repetitions with controlled angular variation. Each melon was captured in five repetitions. For each repetition, the smartphone was rotated around the fruit by approximately ±15° relative to the initial shooting position. This approach introduced moderate viewpoint variation while maintaining visual consistency, thereby enhancing the robustness of the dataset for machine learning applications.
4. Advantages of the acquisition protocol. Natural acquisition conditions: Images were captured while the melon remained in its natural growing position, preserving realistic visual features such as stem attachment and surrounding foliage. Consistent object scale: Maintaining a fixed distance ensured comparable fruit size across images. Controlled visual diversity: Limited angular variation provided sufficient diversity without drastic changes in appearance. Improved dataset robustness: The combination of distance control, angle variation, and repeated captures contributed to a high-quality dataset suitable for image analysis and classification tasks.

2.3 Comparator methods and experimental scenarios

In this study, different preprocessing strategies were applied according to the classification method in order to ensure a fair and meaningful comparison. For the Support Vector Machine (SVM) and standard Neural Network (NN) classifiers, image data were first transformed using Principal Component Analysis (PCA). The success of the proposed approach lies in the PCA-based main feature selection module (MFSM), which effectively suppresses noise, mitigates overfitting, and significantly improves classification accuracy in few-shot fine-grained image classification tasks. PCA was employed to reduce data dimensionality while retaining the most informative features. The number of principal components was limited to 100, capturing the dominant variance in the dataset while maintaining computational efficiency. These PCA-based features were subsequently used as inputs for the SVM and NN classifiers.

In contrast, the Convolutional Neural Network (CNN) model was trained directly on the original image data without PCA-based feature reduction. This design choice is motivated by the inherent capability of CNN architectures to automatically learn hierarchical spatial features through convolutional operations, rendering manual feature extraction unnecessary. Based on these considerations, the experimental evaluation was conducted under the following classification scenarios: (i) PCA combined with SVM (PCA–SVM), (ii) PCA combined with a Neural Network (PCA–NN), (iii) CNN without data augmentation, and (iv) CNN with data augmentation. To enhance generalization performance and mitigate overfitting in the CNN-based experiments, data augmentation techniques were applied during training. The augmentation process included random rotations of up to 20°, horizontal and vertical shifts of up to 10% of the image dimensions, zooming of up to 20%, and horizontal flipping. Data augmentation was performed online using an image data generator, such that transformations were applied dynamically at each training iteration. Consequently, the size of the original training dataset (150 images) remained unchanged, while the model was exposed to diverse transformed versions of the same images across training epochs. For instance, training over 50 epochs effectively resulted in exposure to approximately 7,500 augmented samples (150 images × 50 epochs), although these samples were not explicitly stored. This on-the-fly augmentation strategy improves model robustness by simulating variations in viewpoint and appearance, and is widely adopted as an efficient alternative to offline augmentation in deep learning–based image classification tasks.

3 Research methodology

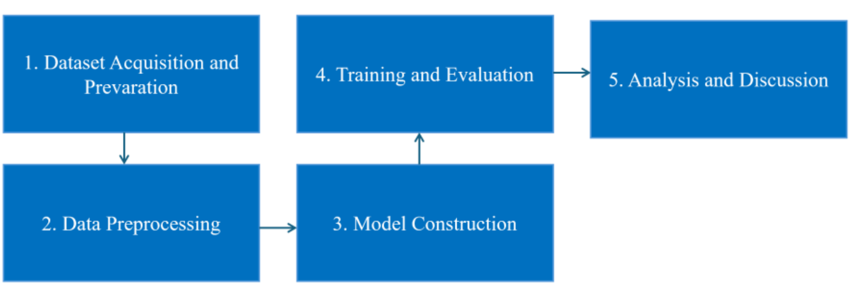

Fig. 3. Five main stages of our research methodology.

The research methodology follows a systematic workflow designed to ensure accurate and reliable classification of melon ripeness (Figure 3). It begins with Dataset Acquisition and Preparation, where image data are collected and organized. This is followed by Data Preprocessing, which standardizes and transforms the data to ensure its suitability for modeling. The third stage, Model Construction, involves building several machine learning and deep learning architectures to perform the classification task. Next, the Training and Evaluation phase assesses model performance using relevant metrics. Finally, Analysis and Discussion provide deeper insights into the results, highlight performance differences between models, and draw conclusions about the effectiveness of the proposed approach. For detail of each process description as follows.

1. Dataset Acquisition and Preparation - The dataset consists of images of Golden Alisha melons collected at four maturity levels (47, 53, 60, and 67 DAP). Images were captured under controlled distance and lighting using a smartphone, generating 230 samples. All images were resized to 516×387 pixels to optimize computation and prepare them for model development.
2. Data Preprocessing - Preprocessing involves normalizing and reducing image dimensions using PCA for traditional models such as SVM and neural networks. For CNN-based approaches, images were processed directly without PCA. Data augmentation (including rotation, shifting, zooming, and flipping) was applied in the augmented CNN scenario to increase data diversity and improve generalization.
3. Model Development - Four models were developed: PCA + SVM, PCA + Neural Network, CNN without augmentation, and CNN with augmentation. The CNN models automatically extracted spatial features, while augmentation enriched the training dataset by generating varied image versions on the fly.
4. Training and Evaluation - All models were trained for 50 epochs and evaluated using accuracy, precision, recall, and F1-score. Confusion matrices were analyzed to identify misclassification patterns. The CNN with augmentation achieved the highest performance, showing improved robustness and accuracy over other methods.
5. Analysis and Discussion - The findings show that PCA-based models lacked sufficient spatial information, reducing their effectiveness. CNNs performed significantly better, and augmentation further enhanced performance by increasing variability. Overall, CNN with augmentation provided the most reliable classification results for melon ripeness.

4 Experiment scenarios

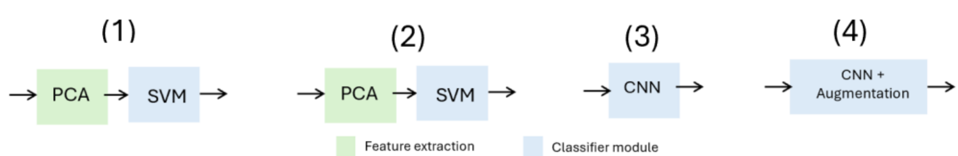

Fig. 4. Four experimental scenarios evaluated in this study

This study was conducted under four distinct experimental scenarios (Figure 4) to evaluate the effectiveness of various classification approaches for melon fruit ripeness classification based on image data. All models across the four scenarios were trained using a uniform number of 50 epochs, allowing for consistent comparison of learning performance across different method.

4.1 PCA with Support Vector Machine (SVM)

In this scenario, image data were first processed using Principal Component Analysis (PCA) to reduce dimensionality and extract the most relevant features. A total of 100 principal components were retained. These reduced-dimension features were then used to train a

Support Vector Machine (SVM) classifier. In this experiment, a Support Vector Machine (SVM) classifier was applied to image data that had been pre-processed through normalization and Principal Component Analysis (PCA). All images size were 516 × 387 pixels, flattened, and standardized using Standard Scaler. Standard Scaler transforms each feature in the dataset so that it has a mean of 0 and a standard deviation of 1. This process is known as standardization. PCA was then employed to reduce the feature dimensionality to 100 principal components. The SVM model used an RBF kernel with parameters C=1 and gamma='scale'.

4.2 PCA with Neural Network (NNs)

Similar to the first scenario, PCA was applied to extract 100 principal components from the image data. The resulting features were then fed into a fully connected feedforward neural network for classification. In this experiment, a fully connected feedforward neural network (NN) was used. Data as an input had been resized, standardized, and reduced to 100 dimensions using PCA. The network architecture consists of an input layer with 100 units, followed by two hidden layers with 128 and 64 neurons respectively, both using ReLU activation functions, and an output layer with 4 neurons using soft max activation to represent the four maturity classes (DAP47, DAP53, DAP60, DAP67). The model was trained using the Adam optimizer and categorical cross entropy loss with a batch size of 8.

4.3 Convolutional Neural Network (CNN) without augmentation

In this scenario, a Convolutional Neural Network (CNN) was trained directly on the raw image data without applying PCA or data augmentation. This setup allowed the CNN to learn spatial hierarchies and extract features automatically from the original images. In this experiment, the CNN architecture consists of three convolutional layers with increasing filter sizes: 32, 64, and 128 filters respectively, each using 3 × 3 kernels and ReLU activation, followed by max pooling layers with a pool size of 2 × 2 to reduce spatial dimensions. After the convolutional blocks, the feature maps are flattened and passed to a dense layer with 128 neurons using ReLU activation, followed by a dropout layer with a rate of 0.3 to prevent overfitting. The output layer uses soft max activation with a number of neurons equal to the number of classes (4) for multi-class classification. The model was compiled with the Adam optimizer and categorical cross entropy loss, and trained for up to 10 epochs using a batch size of 16, with early stopping applied to avoid overfitting and restore the best model based on validation accuracy.

4.4 CNN with augmentation

This final scenario extended the third by incorporating image data augmentation techniques during training using the Image Data Generator class. The augmentations included random rotation, width and height shifts, zooming, and horizontal flipping. These transformations were applied in real-time to enrich the training data and improve model generalization.

5 Results

This section presents the experimental results of the proposed image-based melon ripeness classification framework. The evaluation focuses on two main aspects: (i) the characteristics of the constructed image dataset and (ii) the comparative performance of different classification models under multiple experimental scenarios. The results are reported using standard evaluation metrics, including precision, recall, F1-score, and overall accuracy, to provide a comprehensive assessment of each approach. Through this analysis, the effectiveness of convolutional neural networks and the impact of data augmentation on classification performance are systematically examined.

5.1 New dataset acquisition

As mentioned before, the main contribution of this study is the development of a new melon ripeness dataset that includes both image and acoustic data collected directly from real cultivation environments. The data acquisition was carried out at the Agrotechno Park (ATP) Cikarawang greenhouse, located within the IPB University Experimental Farm in Bogor, West Java, Indonesia. This location provides controlled yet realistic agricultural conditions, enabling systematic observation of melon growth and ripening stages. While the dataset as a whole consists of both image and acoustic modalities, the present study focuses on the image dataset, which is used for visual-based ripeness classification and model development. The image dataset serves as a fundamental component for future expansion toward multimodal approaches that integrate visual and acoustic information to enhance the accuracy of ripeness detection systems.

The dataset consists of a total of 230 primary images, comprising 150 training images and 80 testing images, which were captured using a smartphone camera with a native resolution of 3096 × 4128 pixels (approximately 12 megapixels). To ensure computational efficiency during model training and inference, all images were subsequently resized to a resolution of 516 × 387 pixels while maintaining the original aspect ratio as closely as possible. This resizing step was performed to reduce memory usage and accelerate the processing time, especially during convolution operations in the CNN model. The dataset is categorized into four distinct fruit maturity classes, which are determined based on the number of days after pollination and referred to as DAP (Figure 5). These classes represent progressive stages of ripeness in the Golden Alisha melon variety. Specifically, the 47 DAP class consists of 60 images, the 53 DAP class includes 55 images, the 60 DAP class contains 60 images, and the 67 DAP class comprises 55 images. This classification reflects different physiological stages of fruit maturity, ranging from early to full ripeness. The number of images in each class is relatively balanced, allowing all classifiers to effectively learn and generalize the distinguishing features associated with each maturity stage.

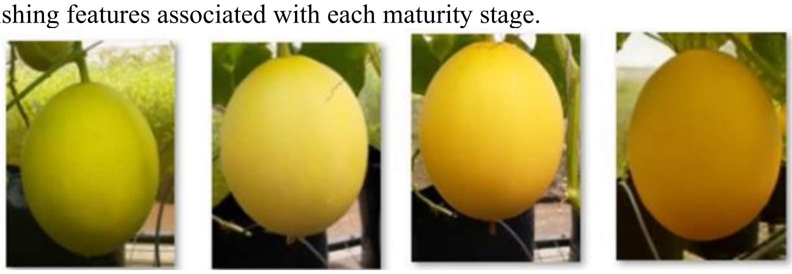

Fig. 5. Sample images for each class (47 DAP, 53 DAP, 60 DAP, 67 DAP).

5.2 Experiment results

The results of the classification experiments are summarized in Table 1 and illustrated visually in Figure 6. These experiments evaluate the performance of four different modelling approaches applied to the melon fruit maturity classification task: (1) PCA combined with Support Vector Machines (PCA + SVM), (2) PCA combined with Neural Networks (PCA + NNs), (3) a Convolutional Neural Network (CNN) trained directly on raw image data, and (4) a CNN with image data augmentation techniques applied during training. The evaluation metrics used for comparison include precision, recall, F1-score, and overall classification accuracy.

Table 1. Experiment result for each scenario

Scenario	Prec.	Rec.	F1-score	Acc.
PCA + SVM	0.73	0.73	0.72	0.72
PCA + NNs	0.65	0.64	0.63	0.64
CNN	0.84	0.78	0.75	0.78
CNN with augmentation	0.89	0.86	0.87	**0.86**

From the results, it is evident that the CNN with augmentation consistently outperforms the other approaches across all metrics. It achieved the highest precision (0.89), recall (0.86), F1-score (0.87), and accuracy (0.86). These results highlight the effectiveness of data augmentation strategies such as rotation, zoom, and horizontal flips in increasing the diversity of the training data and improving the model's generalization capability. The CNN without augmentation also performed strongly, with an accuracy of 0.78, indicating that convolutional models can successfully extract spatial features and patterns from raw images, even without augmented data.

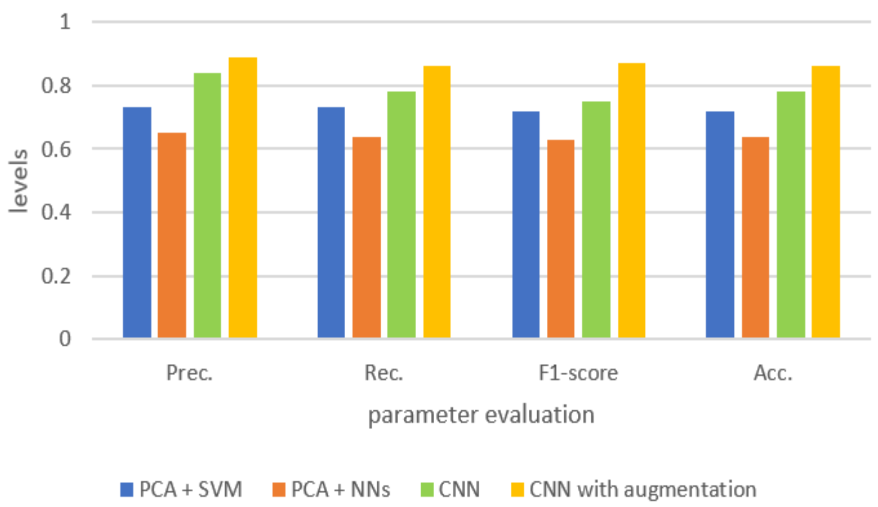

Fig. 6. Experiment results.

In contrast, traditional machine learning models using PCA-reduced features performed less effectively. The PCA + SVM approach yielded a moderate accuracy of 0.72, along with balanced precision and recall values (both 0.73). This suggests that while PCA helps reduce dimensionality and computational complexity, some essential spatial information may be lost in the transformation, limiting the SVM's performance. Meanwhile, the PCA + Neural Network approach produced the lowest performance, with an accuracy of only 0.64. This is likely due to the limited representational power of shallow fully connected networks when working with compressed features that no longer retain the full structural characteristics of the original images. Overall, the results clearly demonstrate that deep learning models, particularly CNNs, are better suited for image-based classification tasks like fruit maturity assessment. The incorporation of data augmentation further enhances the performance, making it the most promising approach among those tested in this study.

6 Discussion

To gain deeper insights into the classification performance of each scenario, confusion matrices (Fig. 6) were analysed for all four experimental configurations. PCA + SVM, PCA + Neural Networks (NNs), CNN, and CNN with data augmentation. These matrices provide a detailed view of how well each model performed in identifying the four melon maturity classes (DAP47, DAP53, DAP60, and DAP67), including the types and frequencies of misclassifications. By analysing the distribution of correct and incorrect predictions, we can identify specific class-level challenges and patterns of confusion, particularly in cases where visual similarity between ripening stages may have affected the model's ability to distinguish between classes accurately.

The confusion matrix for the PCA + SVM scenario shows that DAP47 was the easiest class to predict, with a perfect recall of 1.00 (20/20 correctly classified). In contrast, the DAP60 class had the most prediction errors, with only 11 out of 20 samples correctly classified, and the rest misclassified as DAP53 or DAP67. The confusion between DAP60 and DAP67 suggests a degree of visual similarity between these maturity levels, such as comparable texture or skin tone characteristics that are not well preserved after PCA transformation. Moreover, DAP53 also had notable confusion with DAP47, possibly due to overlapping surface features in mid-stage ripeness.

Table 2. Confusion matrix for each scenario

PCA + SVM

Actual \ Predicted	DAP 47	DAP 53	DAP 60	DAP 67
DAP47	20	0	0	0
DAP53	6	12	2	0
DAP60	0	2	11	7
DAP67	1	3	1	15

(a)

PCA + NNs

Actual \ Predicted	DAP 47	DAP 53	DAP 60	DAP 67
DAP47	14	6	0	0
DAP53	8	8	3	1
DAP60	4	1	12	3
DAP67	2	0	1	17

(b)

CNN without Augmentation

Actual \ Predicted	DAP 47	DAP 53	DAP 60	DAP 67
DAP47	6	14	0	0
DAP53	0	16	4	0
DAP60	0	0	20	0
DAP67	0	0	0	20

(c)

CNN with Augmentation

Actual \ Predicted	DAP 47	DAP 53	DAP 60	DAP 67
DAP47	15	5	0	0
DAP53	0	20	0	0
DAP60	0	4	16	0
DAP67	0	0	2	18

(d)

For PCA + NNs scenario, misclassification was more frequent across all classes. DAP53 and DAP60 were particularly difficult to predict, with their recall values dropping significantly. This is likely due to the limited depth and capacity of the neural network in modelling the complex relationships in compressed PCA features. The visual similarity between DAP53 and DAP6 such as partially developed netting patterns or transitional color changes may cause ambiguity, especially when high-level spatial features are no longer preserved. This highlights the limitation of using shallow networks on reduced-dimension inputs without spatial context.

In the CNN scenario, performance improved markedly compared to PCA-based approaches. The model effectively learned spatial hierarchies in the image data, resulting in higher recall and precision across most classes. However, DAP60 remained the most challenging class, often misclassified as DAP67. This may be due to the visual closeness between the two maturity stages, especially in surface brightness, density of netting, or color tone, which gradually shift and overlap in the later stages of ripeness. Some confusion between DAP53 and DAP47 also remained, likely due to similarities in early skin texture and patterns.

Furthermore, for augmented CNN scenarios, the model demonstrated the most robust and accurate classification. All classes showed improved recall, with DAP60 and DAP53 seeing the greatest gains in prediction accuracy. Although minor misclassifications still occurred, particularly between DAP60 and DAP67, the frequency was lower than in previous scenarios. The data augmentation process likely contributed to better generalization by exposing the model to variations in lighting, orientation, and spatial distortion. This helped the model distinguish between subtly different maturity stages, despite overlapping visual traits such as netting intensity and rind colour gradients.

Across all scenarios, DAP60 consistently emerged as the most difficult class to classify accurately, mainly due to its intermediate ripening stage and close resemblance to both earlier (DAP53) and later (DAP67) classes. Visual similarities such as rind texture, netting, and hue among adjacent maturity levels significantly impacted model performance, particularly in PCA-based models. CNN architectures, especially when combined with data augmentation, proved most effective in mitigating this challenge by learning more abstract and invariant features from raw image data.

Data Diversity - The dataset used in this study comprised 230 images distributed fairly evenly across four ripeness classes (47 DAP, DAP53, DAP60, and DAP67), ensuring a balanced representation but still constituting a relatively small dataset for deep learning applications. Although the controlled image acquisition process maintained consistency in lighting, distance, and angle, it also limited the natural variability typically found in real-world conditions. Consequently, the inherent diversity of the dataset can be considered moderate but not extensive enough to guarantee strong generalization. To address this limitation, data augmentation was applied using real-time transformations such as rotation, translation, zooming, and flipping. This approach effectively expanded the diversity of the training samples, allowing the CNN to learn from a wider range of visual variations and significantly improving its performance—from 78% accuracy without augmentation to 86% with augmentation.

Strengths and Weaknesses of SVM - Support Vector Machine (SVM) demonstrated moderate performance in this study, achieving 72% accuracy when combined with PCA-based feature reduction. Its main strength lies in its robustness with small datasets and its ability to find optimal decision boundaries through margin maximization. However, in image-based tasks, SVM heavily depends on the quality of the manually extracted or reduced features. When used with PCA, much of the spatial structure and contextual information within the original images was lost, resulting in reduced discriminative power. Moreover, SVM struggles with subtle class boundaries and complex visual textures, such as distinguishing between adjacent ripening stages (e.g., DAP60 and DAP67). Therefore, while SVM remains computationally efficient and theoretically sound, it is less suitable for tasks requiring spatial feature learning and high-level abstraction from image data.

Weaknesses of Neural Network (NN) - The fully connected Neural Network (NN) model used in this study produced the lowest performance, with an accuracy of only 64%. This weakness arises mainly from the architecture's limited ability to handle spatial data. Since the input features were first reduced using PCA to 100 components, the network operated on compressed, one-dimensional representations that lacked spatial context. As a result, the NN failed to capture essential visual patterns such as texture gradients or color transitions that indicate ripening stages. Additionally, shallow networks like the one used here have insufficient representational capacity to model complex, nonlinear visual relationships. The lack of mechanisms for spatial invariance, such as translation or rotation tolerance, further reduced its robustness. Hence, traditional feedforward NNs are not well-suited for visual classification tasks that require hierarchical feature learning.

Advantages of CNN Compared to SVM and NN - The Convolutional Neural Network (CNN) outperformed all other models by a significant margin, especially when data augmentation was applied. Unlike SVM and standard NNs, CNNs can automatically learn hierarchical spatial features directly from raw images, eliminating the need for manual feature extraction or dimensionality reduction. This ability allows CNNs to capture essential patterns such as edges, textures, and color distributions that correlate with ripeness stages. The CNN also exhibited greater robustness to variations in lighting, orientation, and background conditions, especially after augmentation. While CNNs require higher computational resources and longer training times, their superior feature extraction and generalization capabilities resulted in the highest accuracy (86%) and best F1-score (0.87). Overall, CNNs, particularly with data augmentation, proved to be the most effective and reliable approach for image-based melon ripeness classification compared to SVM and traditional neural networks.

7 Conclusion

This research conducted a comparative assessment of four classification strategies for identifying the maturity stages of Golden Alisha melons based on image data, namely PCA combined with SVM, PCA combined with Neural Networks, a standard CNN, and a CNN enhanced with data augmentation. The experimental outcomes clearly indicate that deep learning–based approaches, especially Convolutional Neural Networks, consistently surpassed conventional machine learning methods in all evaluated metrics. The best performance was achieved by the CNN model employing data augmentation, attaining an accuracy of 86% along with higher precision, recall, and F1-score values. Augmentation techniques such as image rotation, scaling, and horizontal flipping effectively increased data diversity, thereby improving the model's generalization capability despite the limited dataset size.

Conversely, models relying on PCA exhibited reduced performance, suggesting that dimensionality reduction may discard important spatial information necessary for accurate image-based maturity classification. Analysis of the confusion matrices further showed that the 60 DAP class, representing a transitional ripeness stage, posed the greatest classification challenge due to its visual resemblance to adjacent maturity levels. This finding emphasizes the need for models capable of extracting discriminative spatial features.

In summary, the results confirm that CNN-based models, particularly when combined with data augmentation, provide a reliable and effective solution for image-based fruit maturity assessment. The proposed approach demonstrates strong potential for application in precision agriculture and postharvest quality evaluation of horticultural products.

References

1. Directorate General of Horticulture, Guidelines for Horticultural Agricultural Statistics (OSPHD), Rev. 1, (Directorate General of Horticulture, Ministry of Agriculture of the Republic of Indonesia, Jakarta, 2022)
2. M.M. Abd Zaid, A.A. Mohammed, P. Sumari, Remote sensing image classification using convolutional neural network (CNN) and transfer learning techniques. Journal of Computer Science. **21**, 635–645 (2025). https://doi.org/10.48550/arXiv.2503.02510

3. A. Wicaksono, I.P.E.N. Kencana, I.W. Sumarjaya, Image classification comparison using neural network and support vector machine algorithm with VGG16 as feature extraction method. International Journal of Applied Machine Learning and Computing. **1**, 41-52 (2023). https://doi.org/10.62951/ijamc.v1i3.29
4. M. Wang, B. Zheng, G. Wang, J. Yang, J. Lu, W. Zhang, A principal component analysis–based feature optimization network for few-shot fine-grained image classification. Mathematics. **13**, 1098 (2025). https://doi.org/10.3390/math13071098
5. L. Tiffany, Tabulampot cultivation technique of melon (*Cucumis melo* L.) at Mekarsari Fruit Garden, Cileungsi, West Java, Technical Report, Department of Biology, IPB University, Faculty of Mathematics and Natural Sciences, Bogor (2016)
6. W. Agusta, U. Ahmad, Determination of melon ripeness using acoustic impulse response. Jurnal Keteknikan Pertanian. **4**, 2 (2016).
7. A. Prayoga, H.A. Tawakal, R. Aldiansyah, Pengembangan metode deteksi tingkat kematangan buah melon berdasarkan tekstur kulit buah dengan menggunakan metode Ekstraksi Ciri Statistik dan Support Vector Machine (SVM). Jurnal Teknologi Terpadu. **4**, 24–30 (2018). https://doi.org/10.54914/jtt.v4i1.112
8. R.A. Saputra, D. Puspitasari, T. Baidawi, Melon ripeness detection using support vector machine algorithm based on GLCM feature extraction. Jurnal Infortech. **4**, 2 (2022).
9. A. Alfarizy, F. Fitri, Detection of melon (*Cucumis melo* L.) ripeness using convolutional neural network, unpublished manuscript (2024)
10. M.S. Arham, Deep learning application for melon flower classification, Undergraduate thesis, IPB University, Bogor, Indonesia (2025)