Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1210~1219
e-ISSN: 2548-6861

Implementation of YOLOv11 for Detection and Identification of Strawberry Ripeness Stages

Nilla Mery Handayani [1], Yufis Azhar [2]
Informatics, Universitas Muhammadiyah Malang, Indonesia
nillahandayani@webmail.umm.ac.id [1], yufis@umm.ac.id [2]

Article Info

Article history:

Received 2026-01-17
Revised 2026-02-24
Accepted 2026-04-08

Keyword:

YOLOv11,
Object Detection,
Strawberry Ripeness Stages,
Computer Vision,
Post-Harvest Sorting.

ABSTRACT

This study presents the implementation and evaluation of YOLOv11m, the latest architecture in the YOLO (You Only Look Once) object detection family, for automated multi-stage strawberry ripeness detection. The model was trained on a publicly available Strawberry-DS dataset from Mendeley Data comprising 247 annotated images across six ripeness classes: Green, White, Early-Turning, Turning, Late-Turning, and Red. The dataset was split using stratified sampling into training (70%, 172 images), validation (20%, 49 images), and test (10%, 26 images) subsets. Images were resized to 224×224 pixels and augmented using mosaic, horizontal flipping, HSV adjustment, and random scaling to improve model robustness. Training was performed for 100 epochs using SGD with cosine learning rate scheduling and early stopping. The YOLOv11m model consists of 125 fused layers, 20,034,658 parameters, and 67.7 GFLOPs. Evaluation on the validation set yielded mean precision of 0.748, recall of 0.590, mAP@0.5 of 0.654, and mAP@0.5:0.95 of 0.460. The Red (fully ripe) class achieved the highest performance (mAP@0.5 = 0.889), while transitional classes showed lower scores due to class imbalance and visual similarity. The model achieved an inference speed of 50.1 ms/image (~19 FPS) on a CPU, demonstrating real-time feasibility for industrial deployment. Unlike previous binary or three-class approaches, this study provides fine-grained six-stage ripeness detection with simultaneous object localization, offering a more granular and practically applicable framework for automated post-harvest sorting of strawberries.

I. INTRODUCTION

Strawberries *(Fragaria × ananassa)* represent one of the economically valuable horticultural commodities with continuously increasing demand in both domestic and international markets. This fruit is widely consumed fresh or as raw material in food processing, beverage, and cosmetic industries. The quality of strawberries reaching consumers is greatly determined by their ripeness level, as ripeness affects sensory characteristics such as taste, aroma, texture, and color, all of which play important roles in consumer preference and product competitiveness in the market [1]. Improper ripeness not only reduces organoleptic quality but can also impact economic value decline and increase waste levels in the distribution chain.

In field practice, determining strawberry ripeness stages is generally still done manually through visual observation by workers during harvest or post-harvest sorting processes. This method heavily depends on subjectivity, worker experience, environmental conditions, and physical fatigue factors [2]. Processes that rely heavily on humans tend to be inconsistent, which can lead to irregularities in product quality. This assessment inconsistency becomes more significant at large production scales, where the number of fruits to be sorted per unit time is very high. These conditions lead to high potential for ripeness stage assessment errors. As a result, unripe fruits may be distributed to the market thus reducing product quality, or overripe fruits cause reduced shelf life and increased damage during transportation [3].

Along with the development of science and technology, various automation approaches in fruit quality identification and assessment have been developed to improve sorting process consistency and efficiency. One widely used technology is computer vision, which enables systems to perform image analysis and object recognition automatically [4]. With deep learning support, particularly Convolutional Neural Networks (CNN), the system's capability in extracting visual features becomes much stronger compared to manual feature engineering methods such as color, texture, or morphology extraction. CNNs can learn visual patterns from data directly through the training process, thus producing more complex and accurate feature representations [5].

However, although CNNs can be used to classify images based on ripeness levels, this method has limitations because it only provides class categories without object location information in images. This means CNNs only answer the question "is this image a ripe or unripe strawberry?" but cannot determine where the fruit is located in the image if there are many objects. To overcome this limitation, object detection methods have been developed, which are techniques capable of performing classification while determining object positions through bounding box coordinates [6]. In other words, object detection enables systems not only to know the ripeness category but also to detect the presence of strawberries one by one, so it can be directly applied to automated sorting systems at operational scale.

One of the most widely used object detection methods is YOLO (You Only Look Once). YOLO transforms the object detection process into a single end-to-end regression problem, so the detection process can be performed in one network processing. This makes YOLO much faster compared to region proposal-based methods such as R-CNN and Faster R-CNN [7]. This speed is very important for real-time applications such as agricultural product sorting systems on production lines.

The latest version, YOLOv11, is implemented within the Ultralytics framework and introduces architectural refinements including C3k2 (Cross-Stage Partial Bottleneck with two convolutions) and C2PSA (Cross-Stage Partial with Spatial Attention) modules. These modules improve spatial feature representation and contextual attention while maintaining computational efficiency [8]. The C3k2 module reduces computational cost by splitting the bottleneck into two smaller convolutions, while C2PSA enhances the model's ability to focus on relevant spatial regions. Although YOLOv11 has not yet been extensively documented in peer-reviewed literature, it follows the anchor-free, single-stage detection principles established in recent YOLO generations (YOLOv8, YOLOv9). Compared to YOLOv8m, which serves as the immediate predecessor in the medium-scale variant series, YOLOv11m is reported to provide competitive detection accuracy while maintaining computational efficiency.

Research related to fruit ripeness detection using computer vision and deep learning approaches has been conducted on several commodity types. Saragih and Emanuel [9] applied CNN-based classification for banana ripeness stages using MobileNetV2 and NASNetMobile architectures, achieving up to 96.18% accuracy across four maturity classes. Afonso et al. [10] used deep learning methods including Mask R-CNN for tomato fruit detection and counting in greenhouse environments, demonstrating the applicability of region-based detection networks in agricultural contexts. For strawberries specifically, Tao et al. [11] employed an improved YOLOv5s model (YOLOv5s-BiCE) with a Biformer dual-attention mechanism for three-class strawberry maturity recognition, achieving improved mAP performance over baseline. However, three-class approaches do not capture the full gradation of the ripening continuum relevant to supply chain management. Several prior studies used binary or three-class frameworks, which may limit the representation of subtle transitional stages required for precise post-harvest decision making [12].

Strawberry characteristics differ from many other fruits in that the color transition process occurs gradually and complexly, ranging from white and green through intermediate turning stages to fully red. Precision in distinguishing these stages is critical for supply chain distribution systems that consider shelf life and consumption timing [12]. This motivates the development of a six-class detection framework that more faithfully represents the biological ripening continuum.

Based on the identified challenges, this study focuses on three key aspects. First, it investigates the implementation of YOLOv11m for automatic detection and classification of strawberry ripeness into six categories, namely Green, White, Early-Turning, Turning, Late-Turning, and Red. Second, it evaluates the model’s performance across all ripeness classes to assess its detection consistency and classification reliability. Third, it examines the model’s accuracy and inference efficiency to determine its feasibility for real-time deployment in post-harvest sorting environments.

To address these limitations, this study aims to implement YOLOv11m as an object detection model to identify strawberry ripeness stages automatically in six categories. Model performance is evaluated using standard object detection metrics, including precision, recall, mAP@0.5, mAP@0.5–0.95, and confusion matrix analysis. In addition, inference speed is analyzed to assess the feasibility of real-time implementation.

The main contributions of this study are as follows: (1) the application of multi-stage detection with six granular ripeness stages more suitable for post-harvest industry needs than prior binary or three-class approaches; (2 the implementation of YOLOv11m incorporating C3k2 and

C2PSA modules, which are designed to enhance feature aggregation and computational efficiency; and (3) the integration of simultaneous multi-object localization and classification within a single image frame, supporting automated sorting systems.

II. METHODS

The research stages can be seen in Figure 1.

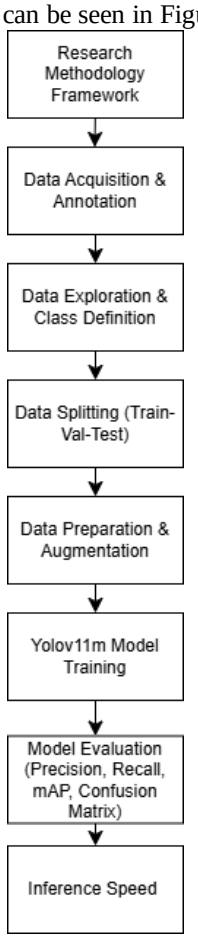

Figure 1. Research Methods Stages

The research stages are illustrated in Figure 1, which presents the proposed experimental framework for multi-stage strawberry ripeness detection. The pipeline includes dataset acquisition and annotation, preprocessing and augmentation, YOLOv11m model training, validation using precision, recall, mAP50, and mAP50-95 metrics, and visualization of detection outputs to evaluate real-time sorting feasibility.

A. Research Methodology Framework

The study follows a structured experimental framework consisting of several sequential stages. First, the research problem was identified based on the limitations of manual strawberry ripeness assessment in post-harvest environments, particularly subjectivity and inconsistency.

Second, a publicly available strawberry dataset from Mendeley Data was selected, containing six annotated ripeness classes. Exploratory analysis was conducted to examine dataset characteristics, including lighting variation, background complexity, fruit orientation, occlusion, and class distribution. Each ripeness stage was explicitly defined based on visible red surface percentage to reduce classification ambiguity.

Next, the dataset was divided using stratified sampling into training (70%), validation (20%), and testing (10%) sets to prevent class imbalance bias. Data preprocessing included image resizing to 224×224 pixels, followed by augmentation techniques such as mosaic augmentation, horizontal flipping, HSV adjustment, and random scaling to improve model robustness.

The YOLOv11m architecture was then selected considering its balance between detection accuracy and computational efficiency. The model was trained for 100 epochs using SGD optimizer with cosine learning rate scheduling and early stopping to prevent overfitting.

Model performance was evaluated using precision, recall, mAP50, mAP50–95, and confusion matrix analysis. Furthermore, error analysis was conducted to identify misclassification patterns among transitional ripeness stages. Finally, inference speed testing was performed to assess real-time feasibility, and computational complexity analysis was carried out to evaluate industrial deployment potential.

B. Data Acquisition & Annotation

This study utilizes a publicly available strawberry dataset obtained from Mendeley Data titled "Strawberry-DS" [13]. The dataset was selected because it provides standardized bounding box annotations in YOLO format, fully compatible with the Ultralytics framework. It includes six distinct ripeness class labels corresponding to the biological progression of strawberry maturation. In addition, the dataset contains natural environmental variations, including different lighting conditions (indoor artificial, outdoor natural, and diffuse illumination), diverse backgrounds (wooden surfaces, sorting trays, and natural foliage), varying fruit orientations and sizes, and partial occlusion scenarios representative of real post-harvest environments.

The total dataset consists of 247 images. Class-level characteristics include significant variation in instance count per class, which is discussed in the data exploration section below. All annotations were verified to follow object detection format with bounding box coordinates (x_center, y_center, width, height) normalized to image dimensions as required by the YOLO framework.

C. Data Exploration & Class Definition

The dataset is annotated into six ripeness stages: Green, White, Early-Turning, Turning, Late-Turning, and Red. Each class represents a specific visual phase in the strawberry ripening continuum observed under field

conditions. The gradual and subtle color transitions between intermediate stages present a challenging detection problem, particularly for distinguishing Early-Turning, Turning, and Late-Turning categories.

Preliminary exploration indicates variation in instance distribution across classes, which may influence model learning behavior. However, no class-weighted loss was applied in this study, and imbalance effects are mitigated through built-in data augmentation techniques during training.

Figure 2. Strawberry Ripeness Stages

Figure 2 shows image illustrates six strawberry ripeness stages, each defined mainly by peel color and surface appearance:

1) *Green*
   The fruit surface is fully green with no visible whitening or redness, indicating an immature stage that is not suitable for consumption or harvest.
2) *White*
   The peel turns pale or whitish, with most of the surface losing its green hue but not yet showing clear pink or red areas; the fruit is still unripe but approaching the ripening phase.
3) *Early-Turning*
   A light pink or slightly orange tint appears on part of the fruit while some areas remain whitish; this is the initial transition from unripe to ripening.

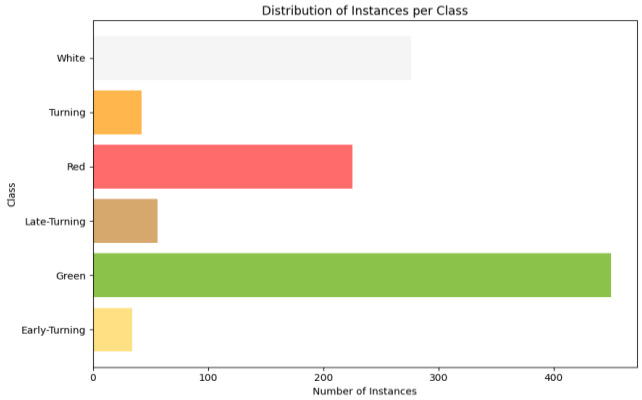

Figure 3. Distribution of Instances per Class

4) *Turning*
   The peel shows a clearer mix of pink and light red over a larger area, though some pale regions are still visible; the fruit is in an intermediate ripening stage.
5) *Late-Turning*
   Most of the surface is red or deep pink with only small patches of lighter color; the fruit is almost fully ripe and close to the harvest stage.
6) *Red*
   The fruit surface is uniformly bright red with a glossy appearance, representing the fully ripe stage that is ideal for harvest and fresh consumption.

Figure 3 distribution of object instances per class across the entire dataset (train, validation, and test combined), totaling 1,083 annotated bounding boxes across 247 images. The distribution reveals significant class imbalance, with Green (450) and White (276) dominating the dataset, while transitional classes Early-Turning (34), Turning (42), and Late-Turning (56) are substantially underrepresented. This imbalance reflects natural harvest batch conditions but poses challenges for model learning of transitional ripeness stages.

D. Data Splitting (Train-Val-Test)

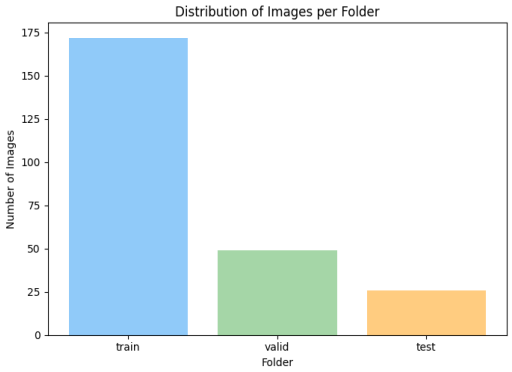

Figure 4. Distribution of Images per Folder

Figure 4 illustrates the distribution of images across the training, validation, and test subsets. This dataset is divided using a 70:20:10 split ratio, resulting in:

1) *Training set: 172 images*
2) *Validation: 49 images*
3) *Test set: 26 images*

Stratified sampling was applied to preserve class distribution across all subsets, ensuring that no subset is disproportionately populated with instances from a single dominant class. This approach prevents evaluation bias and allows for reliable estimation of generalization performance across the full ripeness spectrum.

E. Data Preparation and Augmentation

Before training, images were resized to 224×224 pixels. This resolution was selected as a deliberate compromise between preserving sufficient visual detail for ripeness discrimination and maintaining computational efficiency. While object detection tasks commonly employ higher resolutions (e.g., 640×640), several factors motivated the use of 224×224 in this study:

1) The relatively small dataset size (247 images) means that higher resolutions would not proportionally improve feature learning and would risk overfitting;

2) The primary distinguishing features for strawberry ripeness are color and texture gradients rather than fine spatial detail, both of which are well-preserved at 224×224;
3) This resolution significantly reduces memory consumption and training time on CPU hardware; and
4) It aligns with the real-time deployment objective, as lower resolution reduces inference latency.

This configuration retains essential color and texture information necessary for ripeness stage differentiation while remaining computationally feasible.

To increase data diversity and improve model robustness against real-world variation, augmentation techniques were applied including: translation up to 0.1 times image size; random scaling up to 0.5 of original size; horizontal flipping with probability of 0.5; mosaic augmentation combining four images into one training sample; and random erasing to simulate partial occlusion. Importantly, rotation and extreme color change augmentations were deliberately excluded to preserve the authenticity of color features that are critical identifiers of strawberry ripeness stages. HSV (hue, saturation, value) adjustment within moderate ranges was applied to simulate natural lighting variation without distorting ripeness-indicative color tones.

F. Yolov11m Model Training

YOLOv11 is a recent advancement in the YOLO (You Only Look Once) object detection family, implemented within the Ultralytics framework. The YOLO framework was originally introduced by Redmon et al. [7] as a one-stage detection paradigm for real-time object localization and classification, and has evolved through multiple generations including YOLOv4 [14], YOLOv5, YOLOv8, and YOLOv9 to progressively improve detection accuracy and computational efficiency.

YOLOv11 introduces two key architectural refinements over its predecessor YOLOv8: the C3k2 module and the C2PSA module. The C3k2 (Cross-Stage Partial Bottleneck with two convolutions) module reduces the number of parameters while maintaining feature extraction capability by decomposing the standard bottleneck into two smaller convolution operations, improving gradient flow and computational efficiency. The C2PSA (Cross-Stage Partial with Position-Sensitive Attention) module adds a spatial attention mechanism that allows the network to assign differential importance to different spatial locations within feature maps, which is particularly beneficial for detecting objects at varying positions and scales within a frame. Compared to YOLOv8, YOLOv11 incorporates optimized modules such as C3k2 and C2PSA that enhance feature fusion and spatial focus, which have been associated with improved detection performance on small objects in complex backgrounds [16].

Although YOLOv11 has not yet been extensively discussed in mainstream peer reviewed literature at the time of this writing, it follows the anchor free, single stage detection principles established and validated across the recent YOLO generations. Its implementation within the well-documented Ultralytics framework (v8.3.234) ensures reproducibility and code transparency.

The medium-scale variant, YOLOv11m, was selected for this study based on its balance between detection accuracy and computational efficiency. Compared to the small variant (YOLOv11s), YOLOv11m offers greater capacity for learning complex feature representations across six visually similar ripeness classes. Compared to the large variant (YOLOv11l), YOLOv11m achieves comparable mAP with substantially lower computational cost, which is important for CPU-based deployment in post-harvest sorting environments.

Training was conducted using the Ultralytics framework on a CPU device (Intel Core i5-12500H) with a batch size of 16 and 100 epochs. The 100 epoch configuration is justified by the relatively small dataset size, which allows convergence with fewer iterations. The SGD optimizer was used with an initial learning rate of 0.01, dynamically adjusted using cosine learning rate scheduling that gradually reduces the learning rate from the initial value to near-zero following a cosine annealing curve, enabling fine-grained convergence in later epochs. An early stopping mechanism with patience of 100 epochs was implemented to automatically terminate training if no significant improvement in validation mAP is observed, preventing unnecessary computation while guarding against overfitting.

The composite loss function consists of three components: (i) box loss (weight 7.5) using CIoU for bounding box regression; (ii) classification loss (weight 0.5) for multi-class prediction; and (iii) Distribution Focal Loss (weight 1.5) for refined boundary estimation.

Built in augmentation during training included horizontal flipping (probability 0.5), translation (0.1), mosaic augmentation, RandAugment, and random erasing to increase dataset variability. The model was evaluated on the validation set at every epoch using precision, recall, and mAP metrics to monitor training progress.

G. Model Evaluation (Precision, Recall, Map, Confusion Matrix)

After completing the training process, the YOLOv11 model was evaluated using both the validation and test datasets to assess its capability in detecting and classifying strawberry ripeness stages. The evaluation employed standard object detection metrics, including precision, recall, F1-score, mAP@0.5, and mAP@0.5:0.95. The mAP@0.5 metric measures detection accuracy at an IoU threshold of 0.5, while mAP@0.5:0.95 provides a more rigorous assessment by averaging performance across multiple IoU

thresholds [15]. Thus offering a stronger indication of the model's overall robustness.

Class wise precision and recall were analyzed to determine how effectively the model distinguishes visually similar ripeness categories, particularly transitional stages. Confusion matrix analysis was generated to visualize misclassification patterns, identify systematic error tendencies between adjacent ripeness stages, and quantify the relationship between predicted and ground truth labels. Precision-recall curves and confidence based curves were also generated to identify the optimal confidence threshold for deployment.

H. Inference Speed

To evaluate real-time implementation feasibility, inference speed was measured after training. Average inference time per image (ms) was recorded using the trained YOLOv11m model on CPU hardware. Frames per second (FPS) were calculated using the formula FPS = 1000 / inference_time_ms. The hardware specification (Intel Core i5-12500H without GPU acceleration) was selected to represent a realistic deployment scenario for small-to-medium post-harvest sorting facilities that may not have dedicated GPU infrastructure.

III. RESULTS AND DISCUSSION

In this research, YOLOv11 was implemented to detect and identify strawberry ripeness stages using the Strawberry-DS dataset from Mendeley Data. The evaluation process was conducted comprehensively with quantitative and qualitative visualization so that results can be well understood.

A. Inference Speed Analysis

The inference performance of the YOLOv11 model was evaluated using a CPU-based environment to assess the computational efficiency under practical deployment constraints. The model achieved an average inference time of 50.1 ms per image, corresponding to approximately 19 frames per second (FPS) when executed on an Intel Core i5-12500H CPU without GPU acceleration.

Table I presents the computational complexity of the YOLOv11m model. The network consists of 125 fused layers with 20,034,658 trainable parameters and 67.7 GFLOPs. Despite moderate architectural complexity, the model achieves real-time CPU-based inference at 50.1 ms/image (~19 FPS). For comparison, YOLOv8m achieves approximately 18-22 FPS on similar CPU hardware at 640×640 resolution, suggesting that YOLOv11m at 224×224 achieves comparable throughput with reduced memory overhead.

TABEL I
INFERENCE SPEED AND ACCURACY PERFORMANCE OF YOLOV11

Parameter	Value
Model Architecture	YOLOv11m (Ultralytics v8.3.234)
Task	Object Detection
Number of Classes	6
Total Layers (fused)	125
Total Parameters	20,034,658
GFLOPs	67.7
Input Resolution	224 x 224 pixels
Framework	PyTorch 2.9.1 (CPU)
Hardware	Intel Core i5-12500H (CPU)
Inference Time	50.1 ms/image
Estimated FPS	≈ 19 FPS

In industrial sorting contexts, conveyor belt speeds typically require 10-30 FPS for reliable fruit-by-fruit classification, placing the proposed system within the feasible operational range. For GPU-accelerated deployment, inference times would be expected to decrease to under 5 ms/image, corresponding to over 200 FPS.

The processing speed was calculated using the following equation:

$$FPS = \frac{1000}{T_{inf}}$$

Where T_{inf} denotes the average inference time in milliseconds.

Substituting $T_{inf} = 50.1\ ms$ into the equation yields:

$$FPS = \frac{1000}{50.1} \approx 19.9$$

TABEL II
INFERENCE SPEED AND ACCURACY PERFORMANCE OF YOLOV11

Parameter	Value
Inference Time (ms/image)	$50.1\ ms$
Frames Per Second (FPS)	$\approx 19\ FPS$

As shown in Table II, the YOLOv11 model maintains real-time processing capability while achieving stable detection performance.

The YOLOv11 model was trained for 100 epochs on the Strawberry-DS dataset using an input resolution of 224 x 224. The final validation results on 49 images containing 1.083 annotated strawberry instances are summarized in Table II.

B. Overall Model Performance

Table III demonstrates the overall capability of the trained YOLOv11m model on the Strawberry-DS validation set. The mean precision of 0.748 indicates that approximately 75% of detected strawberries were correctly classified with

the appropriate ripeness stage, demonstrating reasonable control of false positives. However, the mean recall of 0.590 reveals that the model successfully detected only 59% of all ground-truth strawberry instances, suggesting that a substantial portion of fruits particularly those in transitional ripeness stages remained undetected.

TABEL III
OVERALL MODEL PERFORMANCE

Precision	Recall	mAP@0.5	mAP@0.5:0.95
0.748	0.590	0.654	0.460

These metrics collectively demonstrate that the current model configuration is suitable for applications focused on harvesting fully ripe fruit. However, further refinement is necessary to improve recall in transitional classes and enhance bounding box accuracy across the entire ripeness spectrum.

C. Per-Class Performance Analysis

TABEL IV
PER-CLASS PERFORMANCE

Ripeness Stage Class	Instances	Precision	Recall	mAP @0.5	mAP @0.5: 0,95
Green	104	0.757	0.627	0.696	0.357
White	54	0.594	0.542	0.552	0.394
Early-Turning	7	0.944	0.429	0.889	0.728
Turning	10	0.680	0.428	0.632	0.493
Late-Turning	14	0.686	0.571	0.552	0.304
Red	36	0.825	0.944	0.889	0.728

Table IV reveals significant performance variation across ripeness classes. The Red (fully ripe) class demonstrates the strongest performance with precision of 0.825, recall of 0.944, and mAP@0.5 of 0.889. This high performance is attributable to the distinctive uniform bright red color and glossy surface texture that clearly differentiate fully ripe strawberries from all other stages [16]. Similarly, the Green class shows robust metrics with mAP@0.5 of 0.696, benefiting from clear visual differentiation of fully green coloration.

In contrast, the transitional classes (Early-Turning, Turning, Late-Turning, and White) exhibit considerably lower performance, with mAP@0.5 values ranging from 0.552 to 0.632 and notably low recall values between 0.428 and 0.571. The Early-Turning class, despite having the highest precision (0.944), suffers from the lowest recall (0.429), indicating frequent missed detections. This high precision with low recall pattern is consistent with a model that is conservative in predicting this class, only doing so with high confidence in the few cases where it detects the characteristic early-pink coloration, but missing many instances where the color transition is subtle.

These challenges in transitional stages are attributable to subtle color gradients and visual similarities between consecutive ripeness phases, making discrimination difficult even for deep learning models [17]. The limited training instances in some transitional classes (Early-Turning: 7, Turning: 10, Late-Turning: 14) compared to dominant classes (Green: 104, Red: 36) likely contributes to class imbalance issues affecting model generalization.

D. Evaluation Curve Analysis

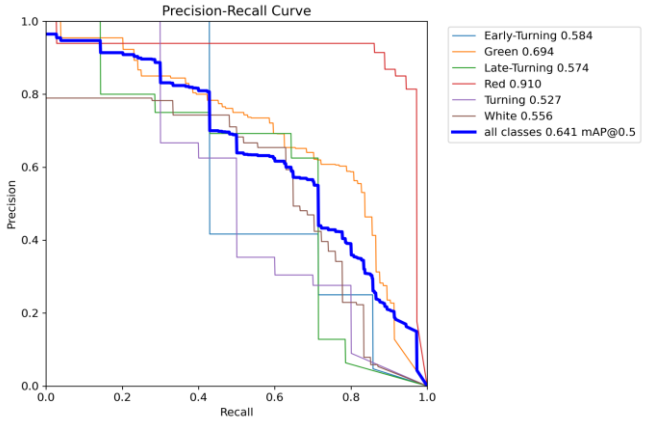

Figure 5. Precision-Recall Curve

Figure 5 presents the Precision-Recall curves for all six ripeness classes. The Red class curve demonstrates a substantially larger area under the curve (AUC) compared to other classes, confirming the model's superior capability in distinguishing fully ripe strawberries. The Green and White classes exhibit relatively high AUC values, while the transitional classes show lower curves with smaller AUC regions, reflecting reduced detection consistency. The overlapping curves among transitional classes suggest the model struggles to establish clear decision boundaries between visually similar ripeness stages.

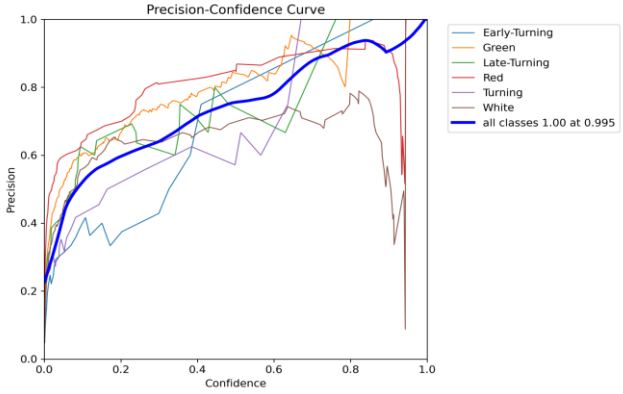

Figure 6. Precision-Confidence Curve

Figure 6 illustrates the relationship between precision, F1-score, and confidence threshold. The optimal detection performance occurs at a confidence threshold of approximately 0.18, where the highest overall F1-score of 0.61 is achieved. This finding has practical implications for deployment: a threshold of 0.18 provides the best precision-recall trade-off for real-time sorting applications. Setting the threshold too high would eliminate many valid detections (reducing recall and causing under-detection of transitional-stage fruits), while too low a threshold would introduce excessive false positives (reducing sorting precision).

E. Confusion Matrix Analysis

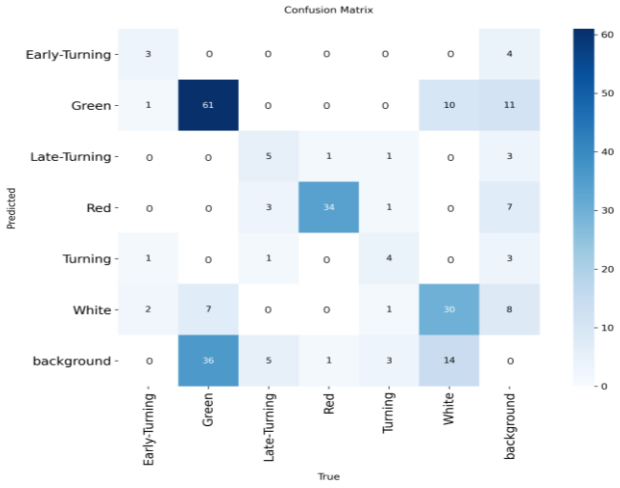

Figure 7. Confusion Matrix

Figure 7 presents the normalized confusion matrix for the YOLOv11m model across all six strawberry ripeness classes. The diagonal elements represent correct classifications, whereas off-diagonal elements indicate misclassification rates between classes.

The confusion matrix reveals several systematic patterns in model performance. The Red class demonstrates the strongest diagonal value, confirming accurate classification of fully ripe strawberries with minimal confusion with other classes. The Green class also shows a relatively strong diagonal, though with notable confusion with White, reflecting the subtle color transition at the initial whitening stage.

Critically, the misclassification analysis reveals that errors are not random but follow the natural ripening sequence. Early-Turning fruits are most frequently misclassified as Turning or White; Turning fruits are confused with Late-Turning; and Late-Turning fruits are occasionally classified as Red. This sequential misclassification pattern confirms that the model has learned the general progression of the ripening continuum, but lacks sufficient resolution to reliably differentiate closely adjacent stages. Notably, misclassifications almost never skip intermediate stages (e.g., Green is not directly misclassified as Red), demonstrating that the learned feature representations capture meaningful aspects of the ripening gradient.

An asymmetry is observable in certain confusion patterns: Late-Turning fruits are more frequently classified as Red than vice versa, suggesting a slight model bias toward predicting more mature ripeness stages when features are ambiguous. This bias may originate from the relative abundance of Red training instances (36) compared to Late-Turning (14), causing the model to favor the Red prediction in ambiguous boundary cases.

The White class shows notable confusion with both Green and Early-Turning classes, which is expected given that white coloration represents a brief transitional phase between the green and pink/red pigmentation phases. The low recall for White (0.542) suggests that many White-stage fruits are being detected but assigned to an adjacent class.

These findings suggest several practical mitigation strategies. In the short term, increasing the confidence threshold for transitional-class predictions and flagging uncertain predictions for human review could improve sorting accuracy. In the longer term, collecting more training instances for Early-Turning, Turning, and Late-Turning classes (recommended minimum: 50+ instances per class) and applying class-weighted loss functions during training could substantially improve transitional-class performance [18].

Additionally, incorporating spectral color features through color space transformations (e.g., HSV, Lab, or even near-infrared spectral features if hardware permits) could provide richer discriminative features for transitional stage differentiation. A hierarchical classification approach that first distinguishes broad categories (Unripe: Green+White; Transitional: Early-Turning+Turning+Late-Turning; Ripe: Red) before applying fine-grained classification could leverage the model's strong extreme-class performance while mitigating transitional-stage challenges.

F. Detection Results Visualization

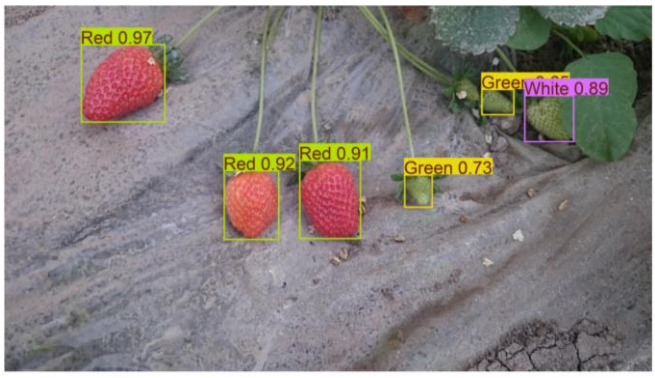

Figure 8. Detection Results Sample 1

The prediction results visualization in Figure 8 shows the model is capable of providing bounding boxes and labels on

strawberry fruits with varying confidence levels for each ripeness stage. The "Red" and "Green" classes tend to be detected with high confidence and good precision. However, in transition classes, incorrect labels or low confidence are often found.

Overall, predictions on real images indicate the model can be used for automatic detection of ripe fruits, but still risks misclassification at intermediate ripeness stages. Namely in the "White", "Turning", and "Early-Turning" classes, misclassification often occurs. This is similar to other research reports that also found high accuracy in ripe categories, but require optimization in intermediate classes for automated sorting applications in the field.

To further demonstrate the model's consistency across multiple samples, Figure 9 presents batch validation predictions under varying environmental conditions, including occlusion, illumination changes, and background complexity.

Figure 9. Validation Detection Results of the YOLOv11m Model

Figure 9 presents batch validation predictions under varying environmental conditions including occlusion, illumination changes, and background complexity. The model demonstrates reasonable robustness to lighting variation and background changes for the clearly distinguishable classes (Red, Green), which is attributable to the HSV augmentation and mosaic augmentation applied during training. For transitional classes, performance degrades more noticeably under challenging illumination conditions, suggesting that additional spectral preprocessing or illumination normalization could improve robustness in deployed environments.

IV. CONCLUSION

This research demonstrates that YOLOv11m implementation is capable of detecting and identifying strawberry ripeness stages with promising performance on the Strawberry-DS dataset from Mendeley Data. The model achieved mean precision of 0.748, recall of 0.590, mAP@0.5 of 0.654, and mAP@0.5:0.95 of 0.460 on the validation set, with inference speed of 50.1 ms/image (~19 FPS) on CPU, confirming real-time feasibility for post-harvest sorting applications.

The model delivers excellent results for the Red (fully ripe) class, achieving precision of 0.825, recall of 0.944, and mAP@0.5 of 0.889, making it highly reliable for identifying harvest-ready strawberries. The Green (unripe) class also demonstrates strong performance with mAP@0.5 of 0.696. These results are attributable to the distinct and contrasting visual characteristics that differentiate extreme ripeness stages.

However, transitional classes (White, Early-Turning, Turning, and Late-Turning) exhibit considerably lower performance, with recall values ranging from 0.428 to 0.571 and mAP@0.5 between 0.552 and 0.632. The confusion matrix reveals frequent misclassifications between adjacent ripeness stages, reflecting the inherent visual ambiguity and gradual color transitions during the ripening process. Additionally, the substantial performance drop from mAP@0.5 (0.654) to mAP@0.5:0.95 (0.460) indicates that bounding box localization accuracy requires improvement, particularly in cases involving fruit occlusion and clustering.

Despite these limitations, the evaluation based on quantitative metrics and prediction visualizations confirms that this system has significant potential for deployment in automated sorting systems and real-time field monitoring for ripe fruit collection. These improvements are expected to enhance model robustness and enable more comprehensive ripeness stage classification for advanced agricultural automation systems.

REFERENCES

[1] F. Giampieri, S. Tulipani, J. M. Alvarez-Suarez, J. L. Quiles, B. Mezzetti, and M. Battino, "The strawberry: Composition, nutritional quality, and impact on human health," *Nutrition*, vol. 28, no. 1, pp. 9-19, Jan. 2012.

[2] J. K. Brecht, M. A. Ritenour, N. F. Haard, and G. W. Chism, "Postharvest physiology of vegetables," in Handbook of Vegetable Science and Technology, D. K. Salunkhe and S. S. Kadam, Eds. New York, NY: Marcel Dekker, 1998, pp. 97–158.

[3] S. J. Kays, Postharvest Physiology of Perishable Plant Products. New York, NY: Van Nostrand Reinhold, 1991.

[4] K. Tanaka and Y. Sato, "Machine vision applications in precision agriculture: A review," Agricultural Systems, vol. 189, pp. 103–121, 2021.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[6] X. Wang and D. Zhang, "Convolutional neural networks for fruit maturity classification: Recent advances and applications," Computers and Electronics in Agriculture, vol. 175, p. 105547, Aug. 2020.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[8] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO11," GitHub repository, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[9] R. E. Saragih and A. W. R. Emanuel, "Banana ripeness classification based on deep learning using convolutional neural network," in 2021 3rd East Indonesia Conference on Computer and Information

Technology (EIConCIT), Surabaya, Indonesia, Apr. 2021, pp. 85–89.
[10] M. Afonso, H. Fonteijn, F. S. Fiorentin, D. Lensink, M. Mooij, N. Faber, G. Polder, and R. Wehrens, "Tomato fruit detection and counting in greenhouses using deep learning," Frontiers in Plant Science, vol. 11, p. 571299, Oct. 2020.
[11] Z. Tao, K. Li, Y. Rao, W. Li, and J. Zhu, "Strawberry maturity recognition based on improved YOLOv5," Agronomy, vol. 14, no. 3, p. 460, Feb. 2024.
[12] C. Wang, H. Wang, Q. Han, Z. Zhang, D. Kong, and X. Zou, "Strawberry detection and ripeness classification using YOLOv8+ model and image processing method," Agriculture, vol. 14, no. 5, p. 751, May 2024.
[13] N. El-Bendary and E. Elhariri, "Strawberry-DS: An annotated benchmark strawberry ripeness dataset," Mendeley Data, v.1, 2022.
[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
[15] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in Proc. Int. Conf. Systems, Signals and Image Processing (IWSSIP), 2020, pp. 237–242.
[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems (NIPS), vol. 28, 2015, pp. 91–99.
[17] P. Behera, A. Behera, and A. 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.
[18] Y. Han, C. Wang, H. Luo, H. Wang, Z. Chen, Y. Xia, and L. Yun, "LRDS-YOLO enhances small object detection in UAV aerial images with a lightweight and efficient design," Scientific Reports, vol. 15, art. no. 22627, 2025.