YOLOv8-Based Tomato Ripeness Detection with Color and Texture Features

Manuscript received August 20, 2026; revised September 8, 2026

Luh Putu Ary Sri Tjahyanti
Dept. of Information Technology
Universitas Panji Sakti
Singaraja, Indonesia
ary.tjahyanti@unipas.ac.id

Made Santo Gitakarma
Dept. of Electrical Engineering
Universitas Pendidikan Ganesha
Singaraja, Indonesia
santo@undiksha.ac.id

Abstract— The advancement of precision agriculture requires intelligent and automated solutions to enhance crop monitoring and post-harvest management. Accurate recognition of tomato ripeness is a critical stage in the agricultural supply chain, as misclassification can lead to quality degradation and significant economic losses. This study proposes a deep learning-based approach for tomato ripeness detection by integrating handcrafted color features and Gray Level Co-occurrence Matrix (GLCM)-based texture features with YOLOv8. The dataset comprises tomato images categorized into three final ripeness classes—green, semi-ripe, and fully ripe—obtained by grouping multiple physiological maturity stages into unified model classes to reduce visual ambiguity and collected under diverse lighting and background conditions. Candidate tomato regions are first localized by YOLOv8; HSV and CIELAB color descriptors and GLCM texture descriptors are then extracted from each region and fused with the corresponding deep representation for ripeness classification and confidence refinement. Experimental results based on five-fold cross-validation show that the proposed method achieves a mean Average Precision (mAP@0.5) of 96.7%, compared with 92.5% for standard YOLOv8. The results show improved discrimination between visually similar ripeness stages, particularly under non-uniform illumination. The study demonstrates the potential of domain-informed feature fusion for smart farming, automated sorting, and quality control systems.

Keywords— Color Features, Deep Learning, GLCM Texture, Precision Agriculture, Tomato Ripeness, YOLOv8.

I. INTRODUCTION

The rapid advancement of information technology and artificial intelligence (AI) has driven fundamental transformations across various sectors, including agriculture. In recent years, the paradigm of precision agriculture has increasingly relied on data-driven technologies, computer vision, and deep learning to enhance production efficiency, reduce losses, and ensure consistent product quality [1], [2], [3]. Within this context, tomato (Solanum lycopersicum) is one of the most economically significant horticultural commodities, whose value is highly dependent on accurate harvest timing and post-harvest handling [4].

Tomatoes are widely consumed both globally and nationally due to their nutritional value and extensive use in the food processing industry [5]. However, the tomato supply chain remains vulnerable to inefficiencies caused by inaccurate ripeness assessment. Fruit maturity directly influences taste, nutritional content, shelf life, and market price [6]. Harvesting tomatoes too early often results in inferior sensory quality and shortened shelf life, whereas delayed harvesting increases the risk of mechanical damage, spoilage, and post-harvest decay [7]. According to FAO statistics, global tomato production reached approximately 192.3 million tons in 2023, while post-harvest losses in developing countries may reach 30–40%, largely due to improper ripeness classification [5], [6]. In Indonesia, distribution-stage losses are estimated at around 33% annually, underscoring the urgent need for reliable and objective ripeness recognition systems.

At present, tomato ripeness assessment in the field is predominantly performed through manual visual inspection by farmers or field workers. Although simple, this approach is inherently subjective, highly dependent on individual experience, and inconsistent when applied at large production scales [8], [9]. Such limitations motivate the development of automated vision-based systems capable of delivering fast, accurate, and consistent ripeness recognition, thereby improving decision-making across harvesting, sorting, and distribution processes.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have enabled significant progress in object detection and visual classification tasks. Modern detection frameworks such as You Only Look Once (YOLO) have demonstrated strong performance in agricultural applications, including pest detection, plant disease identification, and fruit recognition [10], [11], [12]. Among these, YOLOv8 represents the latest generation, offering improved architectural efficiency, faster inference, and competitive accuracy compared to previous versions and alternative detectors [13].

Despite these advantages, accurately distinguishing tomato ripeness stages remains challenging. Differences between maturity levels are often characterized by subtle variations in surface color gradation and skin texture, which are further affected by non-uniform illumination, complex backgrounds, and natural field conditions [14], [15]. Under such circumstances, models that rely solely on CNN-learned features may struggle to capture fine-grained visual cues, particularly for intermediate ripeness stages. Existing YOLO-based studies on fruit ripeness detection, including recent YOLOv8-based approaches largely depend on end-to-end deep features without explicitly incorporating domain-specific descriptors such as color statistics or texture characteristics [15]. Meanwhile, many tomato ripeness studies continue to employ traditional image processing or conventional machine learning techniques, which offer interpretability but lack robustness under real-world variability. This situation reveals a methodological gap between handcrafted feature-based approaches and state-of-the-art deep learning detectors.

 EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

To address this gap, this study proposes a hybrid tomato ripeness recognition framework that explicitly integrates handcrafted color features and Gray Level Co-occurrence Matrix (GLCM)-based texture features into the YOLOv8 detection pipeline. Color features are biologically relevant indicators of tomato maturity, reflecting pigment transitions from green to yellow, orange, and red during ripening [16]. Complementarily, GLCM-based texture features capture spatial relationships among pixels, representing surface smoothness or heterogeneity that correlates with physiological changes during ripening [17]. By fusing these domain-informed features with CNN-learned representations, the proposed approach aims to enrich feature diversity and improve discrimination between visually similar ripeness stages under varying environmental conditions.

Beyond its methodological contribution, this research is particularly relevant to the Indonesian agricultural context. As an agrarian country, Indonesia continues to face challenges related to post-harvest losses, quality inconsistency, and limited adoption of intelligent sorting technologies. An accurate and robust tomato ripeness recognition system has the potential to support automated sorting at both smallholder and industrial scales, enhance supply chain efficiency, and reduce food loss, thereby strengthening the competitiveness of domestic horticultural products.

Accordingly, the objective of this study is to develop and evaluate a YOLOv8-based tomato ripeness detection model augmented with explicit color and GLCM texture features. Model performance is assessed using accuracy and mean Average Precision (mAP) metrics and is systematically compared against the standard YOLOv8 baseline. The key novelty of this work lies in the explicit hybridization of handcrafted color and texture descriptors with a modern YOLOv8 architecture for tomato ripeness detection, providing empirical evidence that domain-informed feature integration can advance detection accuracy and robustness beyond purely CNN-based approaches in real agricultural environments.

II. Related Works

A. Smart Agriculture Technology and Fruit Ripeness Recognition

The rapid adoption of information technology in agriculture has been primarily motivated by the need to improve productivity, reduce post-harvest losses, and ensure consistent crop quality. One of the most critical challenges in horticulture is the accurate assessment of fruit ripeness, as it directly affects market value, shelf life, and consumer acceptance [18]. In practice, tomato ripeness evaluation is still largely conducted through manual visual inspection by farmers, which is inherently subjective and often inconsistent, particularly when applied at large production scales or under varying field conditions [19], [20]. These limitations have driven increasing research interest in automated vision-based ripeness recognition systems.

Previous studies on fruits such as mangoes, bananas, and apples have demonstrated that automated classification methods can significantly reduce post-harvest losses, with reported improvements of up to 20–30% in sorting accuracy and handling efficiency. Such findings highlight the potential of intelligent recognition systems to support harvesting and post-harvest decision-making. However, compared to other fruits, tomato ripeness recognition poses additional challenges due to the gradual and visually subtle transition between maturity stages. This motivates further research focused specifically on tomatoes, particularly under real agricultural conditions where lighting variability and background complexity are unavoidable.

B. Extraction of Color and Texture Features in Maturity Classification

Color is one of the most informative biological indicators of tomato ripeness, reflecting pigment transformations from green to yellow, orange, and red during maturation. Consequently, numerous studies have utilized color representations such as RGB, HSV, and Lab* color spaces to capture ripeness-related variations effectively [21]. While color features provide strong global cues, they are often sensitive to illumination changes and may be insufficient for distinguishing intermediate ripeness stages.

To complement color information, texture analysis has been widely explored in fruit maturity classification. The Gray Level Co-occurrence Matrix (GLCM) is a popular statistical texture descriptor that captures spatial relationships among pixel intensities, enabling the extraction of features such as contrast, correlation, energy, and homogeneity [22], [23]. Several studies have reported that combining color and texture features improves classification accuracy compared to using color alone. Nevertheless, most of these approaches rely on conventional machine learning classifiers, such as Support Vector Machines (SVM) or k-Nearest Neighbors (KNN) [24]. Although interpretable, these models often lack robustness when deployed in complex field environments characterized by non-uniform illumination, shadows, and cluttered backgrounds. This limitation suggests that handcrafted features alone may not be sufficient for real-world tomato ripeness recognition, motivating the integration of such domain-specific descriptors with more powerful deep learning architectures.

C. Deep Learning for Fruit Identification and YOLOv8-Based Approaches

The emergence of deep learning has fundamentally advanced object recognition and classification tasks in agriculture. Convolutional Neural Networks (CNNs) have demonstrated strong capability in learning hierarchical and discriminative representations directly from fruit images, reducing reliance on manual feature engineering [25]. Among object detection frameworks, the You Only Look Once (YOLO) family has gained widespread adoption due to its favorable trade-off between real-time inference speed and detection accuracy.

YOLOv8, as the latest generation of the YOLO series, introduces architectural refinements that improve generalization, computational efficiency, and detection performance compared to earlier versions [26]. Recent studies have successfully applied YOLO-based models to fruit detection and ripeness classification tasks involving bananas, apples, mangoes, and other horticultural products [27], [28]. However, the majority of these implementations rely exclusively on CNN-learned features and treat the detection pipeline as a fully end-to-end process.

In the specific context of tomato ripeness detection, only limited YOLOv8-based studies have been reported. Notably, Yang et al. [15] proposed an improved YOLOv8 architecture for tomato ripeness recognition by modifying network components and training strategies. While achieving promising

 EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

results, their approach relies solely on deep features extracted by the CNN, without explicit incorporation of domain-informed handcrafted descriptors such as color statistics or texture features. As a result, fine-grained ripeness distinctions particularly at transitional stages remain challenging under varying environmental conditions.

D. Comparative Positioning of This Work

To clarify the novelty and positioning of the present study, closely related YOLO-based works were compared qualitatively. Yang et al. [15] optimized YOLOv8 for tomato ripeness recognition while relying on CNN-learned features, whereas other YOLO-based fruit studies [27], [28] primarily used end-to-end CNN representations for real-time detection. In contrast, the present work combines ROI-level deep features with explicit HSV/CIELAB color descriptors and GLCM texture descriptors under heterogeneous field and public-image conditions.

Existing YOLOv8-based tomato ripeness studies therefore primarily emphasize architectural or training improvements while relying on CNN-derived representations. The present study instead integrates biologically meaningful color descriptors and GLCM-based texture features into the YOLOv8 pipeline. This hybrid strategy bridges traditional feature-based image analysis and modern object detection, with the aim of improving discrimination of visually subtle ripeness stages under real agricultural conditions.

III. Material and Methods

A. Research Design

This study uses an experimental deep learning approach with a comparative testing design. The proposed model integrates handcrafted color descriptors and Gray Level Co-occurrence Matrix (GLCM) texture features into a YOLOv8-based ripeness detection pipeline. Its performance is compared with standard YOLOv8 under the same dataset split, input resolution, and training protocol to assess the contribution of feature integration. The experimental design focuses on detection/classification performance and robustness under diverse environmental conditions.

B. Dataset

The dataset consists of 4,500 tomato images representing three final ripeness classes used for model training and evaluation: green, semi-ripe, and fully ripe. Images were collected through field photography and supplemented with the publicly available real-world in-field tomato image dataset reported by Khatun et al. [8] (Data in Brief, 2023, doi: 10.1016/j.dib.2023.109688). All images were standardized to 640 × 640 pixels to comply with the YOLOv8 input requirements.

During the annotation process, four physiological ripeness stages were initially identified based on horticultural standards, namely Immature Green, Color Change Period, Firm and Ripe Period, and Fully Ripe Period. To reduce visual ambiguity and class overlap during model learning, these physiological stages were subsequently mapped into three unified model classes. Specifically, Immature Green was mapped to the green class, Color Change Period to the semi-ripe class, while Firm and Ripe Period and Fully Ripe Period were merged into a single fully ripe class. This merging strategy was adopted because both stages exhibit dominant red coloration with only subtle visual differences that are difficult to distinguish reliably under varying field illumination.

Annotation was performed manually by agricultural experts following established horticultural ripeness standards to ensure labeling consistency. In total, 4,500 images were obtained and distributed evenly across the three model classes: 1,500 green (33.3%), 1,500 semi-ripe (33.3%), and 1,500 fully ripe (33.3%). The initial partition consisted of 70% training (3,150 images), 15% validation (675 images), and 15% independent testing (675 images). Multiple tomato varieties, including cherry, beef, and local Indonesian cultivars, were included to enhance model generalizability across different fruit sizes and appearances. Partitioning was performed before augmentation using source/acquisition sequence as the grouping unit, so images originating from the same tomato or near-duplicate acquisition sequence were retained within a single subset to minimize data leakage between training and evaluation data.

Table I summarizes the physiological ripeness stages, their corresponding final model class labels, and explicit visual descriptions used during dataset annotation. This mapping provides transparent labeling criteria, ensures consistency across the dataset, and aligns the annotation protocol with the three-class classification framework adopted throughout this study.

TABLE I. TOMATO RIPENESS STAGES, CORRESPONDING LABELS, AND EXPLICIT DESCRIPTIONS USED IN THE DATASET

Physiological Ripeness Stage	Example RGB Image	Label	Explicit Description
Immature Green		Green	The fruit exhibits dense flesh and a green peel, indicating an immature stage suitable for long-haul transport..
Color Change Period		Semi-ripe	The fruit surface transitions from green to pale red, accompanied by increasing sugar content, making it suitable for short-haul distribution.

Physiological Ripeness Stage	Example RGB Image	Label	Explicit Description
Firm and Ripe Period		Fully ripe	Red coloration appears on more than 75% of the peel, indicating readiness for immediate consumption.
Fully Ripe Period		Fully ripe	The fruit shows uniform red coloration and reaches peak sugar levels, making it suitable for local distribution.

C. Data Pre-processing

Data pre-processing in this study was designed to improve model robustness and to support the integration of handcrafted features with deep learning-based detection. The first stage involved image augmentation, including random rotation (±15°), horizontal and vertical flipping, controlled illumination variation, and Gaussian noise injection, to increase data diversity and reduce overfitting.

Next, each tomato region of interest (ROI) was converted from RGB to HSV and CIELAB (L*a*b*) color spaces. For HSV, the mean and standard deviation of the H, S, and V channels were computed, yielding six features. For CIELAB, the mean and standard deviation of the L*, a*, and b* channels were computed, yielding another six features. The resulting color descriptor therefore contained 12 features. These color spaces were selected to represent hue/saturation and perceptually meaningful green-to-red transitions associated with tomato ripeness. All handcrafted features were standardized using z-score normalization, with the mean and standard deviation estimated exclusively from the training portion of each fold and then applied unchanged to the corresponding validation data.

In addition to color information, texture features were extracted using the Gray Level Co-occurrence Matrix (GLCM). For each tomato ROI, contrast, correlation, energy, and homogeneity were calculated at four spatial orientations (0°, 45°, 90°, and 135°) with a fixed pixel distance of one, resulting in a 16-dimensional texture vector. The normalized 12-dimensional color vector and 16-dimensional GLCM vector were concatenated to form a 28-dimensional handcrafted descriptor before fusion with the corresponding deep feature representation.

To address environmental variability in field images, illumination correction was applied using Contrast Limited Adaptive Histogram Equalization (CLAHE) and Retinex-based adjustment, while shadow removal was employed to mitigate the effects of uneven lighting. In addition, background segmentation was experimentally applied to reduce false detections caused by dense foliage; however, all final evaluations reported in this study were conducted without explicit background masking to preserve real-field deployment conditions.

The fusion sequence was implemented after initial object localization. The standard YOLOv8 backbone and neck first generated multi-scale deep features, and the localization branch produced candidate tomato bounding boxes. Each candidate box was then used to crop an ROI from the input image, from which the color and GLCM descriptors were computed. The candidate location was associated with the corresponding detector feature-map region, from which a pooled 256-dimensional CNN representation was obtained. The 28-dimensional handcrafted vector was concatenated with this ROI-associated deep representation and forwarded to the classification/confidence refinement branch. Thus, handcrafted descriptors did not generate the bounding boxes; they were used after candidate localization to refine ripeness-class prediction and confidence.

D. Model Architecture

The proposed architecture retains the standard YOLOv8 backbone and neck for multi-scale representation learning and separates candidate localization from handcrafted-feature refinement. First, the conventional localization branch predicts candidate bounding boxes. Second, each candidate ROI is used to compute the HSV/CIELAB color descriptor and the 16-dimensional GLCM texture descriptor. Third, these normalized handcrafted descriptors are concatenated with the ROI-associated CNN feature vector. The fused representation is passed to the ripeness classification and confidence refinement layers, while bounding-box regression remains in the standard localization branch.

This design resolves the dependency between ROI extraction and feature fusion: the bounding box is available before handcrafted features are computed. The ROI-associated CNN representation contains 256 features and the handcrafted descriptor contains 28 features, producing a 284-dimensional fused vector. After concatenation, a fully connected projection layer maps the 284-dimensional vector to 256 dimensions, followed by a ReLU activation and dropout of 0.2 before the ripeness classification/confidence refinement layer. Bounding-box regression remains in the standard localization branch. Fig. 1 summarizes the revised sequence.

 EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

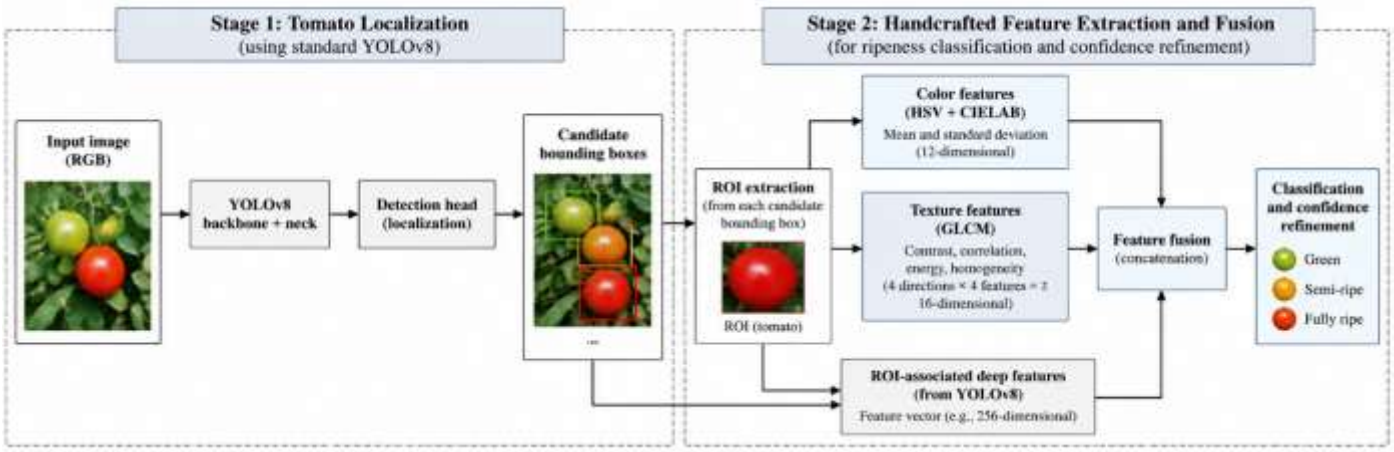


Fig. 1. Revised Two-Stage Architecture for ROI-Based Color and GLCM Feature Fusion with YOLOv8.

E. Model Training

Model training was conducted using the official YOLOv8 implementation in PyTorch on an NVIDIA RTX 3060 GPU with 12 GB VRAM, supported by CUDA 11.8 and cuDNN acceleration. Before augmentation, the dataset was partitioned into 70% training, 15% fixed validation, and 15% independent test subsets. The independent test set was isolated before model development and was not used for hyperparameter tuning or cross-validation.

Five-fold cross-validation was applied only to the 70% training subset. In each fold, four-fifths of the training subset were used to fit the model and one-fifth served as the fold-specific validation portion. Fold-specific mAP@0.5 values were retained for robustness assessment and paired statistical testing. The fixed 15% validation subset was reserved for final configuration/checkpoint confirmation after cross-validation, while the 15% independent test subset remained untouched during model development. This workflow was applied after the initial 70/15/15 partition and before any augmentation, ensuring that augmentation was restricted to the training data.

Training was performed for a maximum of 200 epochs, with early stopping applied when no improvement in validation performance was observed for 20 consecutive epochs. A cosine learning rate scheduling strategy was used to stabilize convergence. On average, each training run required approximately four hours to complete.

For each cross-validation fold, the checkpoint with the highest fold-validation mAP@0.5 was retained. Unless otherwise stated, the performance values reported in Table IV are the mean values across the five cross-validation folds and are not independent-test-set results. The independent test subset was retained as a held-out set and remained excluded from model development, hyperparameter tuning, and cross-validation. Accordingly, Tables IV–VI report the cross-validation results used for the comparative and statistical analyses.

TABLE II. MODEL TRAINING HYPERPARAMETERS

Parameter	Value / Setting	Notes
Framework	PyTorch (YOLOv8 official repository)	CUDA 11.8, cuDNN acceleration
Hardware	NVIDIA RTX 3060 (12 GB VRAM)	~4 hours for 200 epochs
Input resolution	640 × 640 pixels	YOLOv8 default input size
Dataset size	4,500 images	70% train, 15% validation, 15% test
Batch size	32	Optimized for GPU memory
Initial learning rate	0.001	Cosine learning rate scheduling
Optimizer	Adam (modified)	Used instead of default optimizer
Weight decay	1×10^{-5}	Regularization
Dropout	0.2 (classification head only)	Applied after feature fusion
Maximum epochs	200	Early stopping (patience = 20)
Cross-validation	5-fold (training set only)	Validation & test sets fixed
Model selection	Best mAP@0.5	Selected on validation set

F. Performance Evaluation

Model performance was evaluated using standard object detection metrics to ensure a comprehensive and reliable assessment. The primary evaluation metrics were mean Average Precision at IoU thresholds of 0.5 (mAP@0.5) and 0.5:0.95 (mAP@0.5:0.95), which quantify detection accuracy across different levels of localization strictness. These metrics were used as the main indicators for comparing the proposed model with baseline and competing methods.

In addition, classification accuracy, precision, and recall were reported as supporting metrics to provide further insight into class-level performance and misclassification behavior. Confusion matrices were generated for each ripeness class to analyze common error patterns.

The evaluation reported in this manuscript is limited to metrics for which experimental results are provided: mAP@0.5, mAP@0.5:0.95, accuracy, precision, recall, and class-level error analysis. Claims regarding FPS, inference time, CPU/GPU memory usage, Faster R-CNN, and EfficientDet were removed because corresponding experimental results were not available in the current study.

G. Comparative Analysis

The primary quantitative comparison was conducted against the standard YOLOv8 baseline using the same dataset split, input resolution, training schedule, and evaluation protocol. To isolate the contribution of each handcrafted feature family, the ablation design comprised four configurations: standard YOLOv8, YOLOv8 + color, YOLOv8 + GLCM, and YOLOv8 + color + GLCM. The complete ablation results are reported in Table V. The color-only configuration increased mAP@0.5 from 92.5% to 94.5%, while the GLCM-only configuration reached 94.0%; combining both

feature families produced the highest mAP@0.5 of 96.7%, corresponding to a 4.2 percentage-point gain over the baseline.

For statistical comparison, the five fold-level mAP@0.5 scores of the baseline and full hybrid model were treated as paired observations. Mean ± standard deviation and the exact two-sided paired t-test p-value are reported together with the individual fold scores in Table VI, providing both the magnitude and statistical consistency of the observed performance difference.

IV. Results

A. YOLOv8 Model Performance with Color and Texture Feature Integration

Across the five cross-validation folds, the hybrid YOLOv8 model incorporating handcrafted color and GLCM-based texture features achieved a mean mAP@0.5 of 96.7%, compared with 92.5% for the standard YOLOv8 baseline. The 4.2 percentage-point improvement suggests that domain-specific color and texture information provides complementary cues under varying lighting and background conditions.

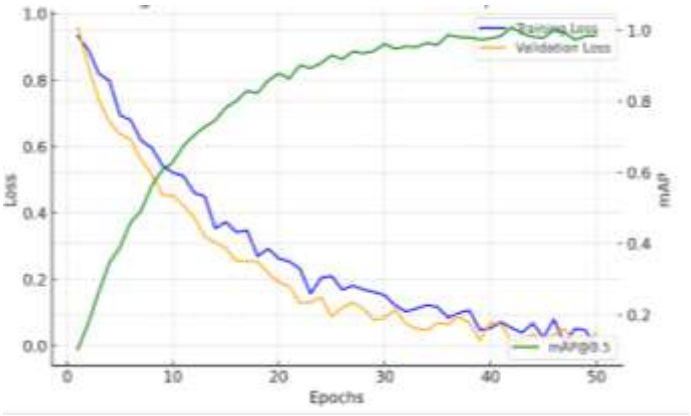


Fig. 2. Training and validation loss with mAP@0.5 performance of YOLOv8 + Color + GLCM model.

Fig. 2 illustrates the training and validation loss curves together with the corresponding mAP@0.5 values of the proposed model. The smooth convergence of the loss functions and the stable increase in detection accuracy across epochs indicate a well-optimized training process and robust model learning behavior.

Under the stricter mAP@0.5:0.95 metric, the cross-validation mean increased from 85.4% for the baseline YOLOv8 to 90.1% for the hybrid model. For the fold-level mAP@0.5 comparison, the baseline achieved 92.5 ± 0.52%, whereas YOLOv8 + Color + GLCM achieved 96.7 ± 1.16%. A two-sided paired t-test across the five folds yielded $t(4) = 12.385$ and $p = 0.000244$, demonstrating a statistically significant and consistent fold-wise improvement for the hybrid model.

The enhanced performance was observed across all three ripeness classes green, semi-ripe, and fully ripe indicating that the proposed feature integration is useful across maturity stages. Table III summarizes the cross-validation mean classification metrics for each class. These values should be distinguished from any final independent-test-set evaluation.

The most pronounced performance improvement was observed in the semi-ripe class, which is visually the most ambiguous stage due to the coexistence of green and reddish regions on the tomato surface. Under natural and non-uniform lighting conditions, this ambiguity frequently causes misclassification in standard YOLOv8 models that primarily rely on color-based convolutional features. As a result, semi-ripe tomatoes are often incorrectly identified as either green or fully ripe.

Table III. Five-Fold Cross-Validation Performance Comparison between Standard YOLOv8 and YOLOv8 with Color and GLCM Features for Three-Class Tomato Ripeness Classification

Ripeness Class	Model	Accuracy (%)	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Green	YOLOv8 Standard	93.1	92.4	91.6	94.0	86.7
	YOLOv8 + Color + GLCM	96.2	95.8	95.1	97.5	90.5
Semi-ripe	YOLOv8 Standard	87.4	86.8	85.2	89.5	83.0
	YOLOv8 + Color + GLCM	94.6	94.1	93.7	96.3	89.2
Fully ripe	YOLOv8 Standard	94.5	93.9	92.8	93.9	86.5
	YOLOv8 + Color + GLCM	97.4	96.8	96.1	98.0	90.6
Average	YOLOv8 Standard	91.7	91.0	89.9	92.5	85.4
	YOLOv8 + Color + GLCM	96.1	95.6	95.0	96.7	90.1

Table IV. Ablation Study of Handcrafted Feature Contributions

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Δ mAP@0.5
Standard YOLOv8	92.5	85.4	—
YOLOv8 + Color	94.5	87.7	+2.0
YOLOv8 + GLCM	94.0	87.1	+1.5
YOLOv8 + Color + GLCM	96.7	90.1	+4.2

Table V. Fold-Level mAP@0.5 Scores and Paired Statistical Test

Fold	Standard YOLOv8	YOLOv8 + Color + GLCM
1	91.8	95.0
2	92.1	96.2
3	92.7	97.7
4	93.0	96.8
5	92.9	97.8
Mean ± SD	92.5 ± 0.52	96.7 ± 1.16

Paired t-test: $t(4) = 12.385$, exact two-sided $p = 0.000244$.

EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

By integrating GLCM-based texture features, the proposed model becomes more sensitive to local surface variations that are not fully captured by convolutional filters alone. Texture descriptors such as contrast and homogeneity provide additional information about pixel distribution and surface irregularities. In particular, semi-ripe tomatoes exhibit higher texture heterogeneity compared to fully ripe tomatoes, which tend to have smoother and more uniform surfaces. This complementary information significantly reduces inter-class confusion in the semi-ripe category.

In addition, the incorporation of color features from HSV and CIELAB provides explicit descriptors of hue, saturation, luminance, and green-to-red chromatic transitions that are strongly associated with tomato maturity [29]. Combined with GLCM texture information, these descriptors enrich the representation available to the refinement classifier and help distinguish visually adjacent ripeness stages.

These findings are consistent with previous studies reporting that hybrid approaches combining deep learning with handcrafted features can improve performance in tasks involving complex visual variability, such as agricultural product classification [29]. While convolutional neural networks are effective at learning global spatial patterns, they are not always optimal for capturing fine-grained local textures. The integration of GLCM introduces domain-relevant information, aligning with the concept of domain-informed deep learning, where expert knowledge enhances data-driven models.

From a practical perspective, improving classification accuracy at the semi-ripe stage is particularly important, as this stage often determines harvesting, sorting, and distribution decisions. Misclassification at this stage can negatively affect product quality, supply chain efficiency, and market value. Therefore, the proposed approach offers clear advantages for real-world applications, especially in automated sorting systems and post-harvest quality control, where reliable ripeness detection directly contributes to reduced losses and improved economic outcomes.

B. Classification Error Analysis

Despite the notable performance gains achieved by the proposed YOLOv8 model with integrated color and GLCM texture features, several misclassification cases remain under sub-optimal environmental conditions, indicating inherent challenges in real-field deployment. The most significant source of error arises from extreme illumination conditions, including both overexposure caused by intense direct sunlight and underexposure in low-light environments. Under these conditions, surface color information becomes saturated or suppressed, reducing the effectiveness of color normalization and causing biased pixel distributions that negatively affect feature extraction. Quantitative analysis shows that misclassifications associated with extreme lighting account for approximately 7.5% of the observed error cases, making illumination variability the dominant limiting factor.

This observation is consistent with prior findings that identify lighting as a critical determinant of performance in outdoor computer vision systems [30]. Beyond illumination, background complexity represents another major contributor to classification errors. Dense foliage, overlapping stems, and hard shadows often introduce texture patterns and color distributions that partially resemble tomato surfaces. In such scenarios, the model may produce false positives or incorrectly assign maturity classes. A recurring failure case involves semi-ripe tomatoes partially occluded by leaves, which are frequently misclassified as green due to the dominance of green spectral components in the surrounding region. This suggests that, although texture features enhance discrimination, the current bounding-box–based detection pipeline is still sensitive to background interference.

These findings indicate that the robustness of hybrid deep learning models is influenced not only by feature representation but also by data quality and environmental variability. While the integration of handcrafted color and texture descriptors strengthens discriminative capacity, it cannot fully compensate for severe visual distortions caused by illumination extremes and occlusion. Therefore, further performance improvements should be pursued through multiple complementary strategies.

First, advanced data augmentation techniques that explicitly simulate extreme lighting, shadow patterns, and partial occlusions should be incorporated to increase model exposure to difficult conditions. Second, the integration of object-level segmentation mechanisms, such as Mask R-CNN or SAM-based segmentation, could better isolate tomato regions from complex backgrounds, reducing false detections and maturity confusion. Third, the adoption of illumination-aware preprocessing, including adaptive shadow removal, Retinex-based normalization, or attention-guided feature weighting, may further stabilize feature extraction. Finally, synthetic data generation using GANs or diffusion-based models offers a promising avenue to enrich training data with rare or hard-to-capture scenarios, thereby improving generalization.

From an application standpoint, this error analysis highlights a key challenge in deploying tomato ripeness recognition systems in real agricultural environments: achieving consistent reliability under uncontrolled conditions. While feature integration substantially enhances performance, addressing environmental variability remains essential for scalable and robust real-world adoption.

C. Comparison with Previous Studies

The findings are consistent with the broader food-vision literature showing that CNN-derived deep features are effective for representing complex visual patterns. Liu et al. [32], for example, reviewed efficient CNN feature extraction for the analysis of complex food matrices and highlighted the discriminative value of learned visual representations. The present results extend this perspective by examining whether explicit domain descriptors can provide complementary information for a fine-grained ripeness task.

Trinh and Nguyen [33] implemented YOLOv5 for real-time pineapple maturity detection, demonstrating the suitability of YOLO-family detectors for fruit maturity recognition. Their approach, however, did not explicitly fuse handcrafted HSV/CIELAB color descriptors and GLCM texture features with the detector. This difference motivates the present hybrid formulation, which combines learned and domain-informed representations.

In contrast to these prior works, the present study integrates color (HSV and CIELAB) and texture (GLCM) descriptors within a YOLOv8-based ROI refinement pipeline. Unlike earlier YOLO-based tomato studies such as Yang et

 EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

al. [15], which rely on CNN-derived features, the proposed method explicitly augments deep representations with handcrafted descriptors intended to capture subtle chromatic and surface-texture differences at transitional ripeness stages.

Moreover, while previous studies typically evaluated hybrid features on controlled datasets or single fruit varieties, this research validates the proposed approach on multiple tomato varieties and heterogeneous field conditions, thereby improving generalizability. Empirically, the proposed model achieves consistent gains across all maturity classes, with the most pronounced improvement observed in the semi-ripe category, an outcome not explicitly addressed in earlier works.

From a methodological perspective, this study advances the literature by demonstrating that multi-dimensional handcrafted feature fusion (color + texture) yields a synergistic effect that surpasses one-dimensional feature enrichment strategies. Rather than duplicating information learned by CNNs, handcrafted features provide complementary cues aligned with horticultural knowledge, supporting the concept of domain-informed deep learning. Consequently, this work offers new empirical evidence that state-of-the-art object detection models such as YOLOv8 can be systematically strengthened through the integration of relevant manual features, bridging the gap between traditional image analysis and modern end-to-end detection frameworks.

V. Conclusion

This study proposed a hybrid tomato ripeness recognition framework that integrates handcrafted color and GLCM-based texture features into the YOLOv8 architecture. Experimental results demonstrate that the proposed approach consistently outperforms the standard YOLOv8 model, achieving 96.7% mAP@0.5 and 90.1% mAP@0.5:0.95. The most notable performance improvement was observed in the semi-ripe (transitional) stage, where texture descriptors such as contrast and homogeneity provided critical discriminative cues that are often difficult to capture using purely convolutional features.

From an academic perspective, these findings reinforce the value of domain-informed deep learning, showing that the combination of handcrafted and learned representations can enhance robustness in agricultural vision tasks. Rather than replacing end-to-end learning, handcrafted features act as complementary descriptors that enrich feature representation under challenging field conditions, such as variable lighting and subtle surface variations.

From a practical standpoint, the proposed model offers a promising solution for smart farming and automated post-harvest systems, enabling more accurate harvest timing, sorting, and quality control. With further optimization, the framework can be deployed on portable or embedded devices to support real-time decision-making in agricultural supply chains. Nevertheless, challenges related to extreme illumination, background complexity, and computational efficiency remain and should be addressed in future work.

Future research will focus on improving environmental robustness through advanced illumination normalization and segmentation strategies, as well as exploring multispectral or hyperspectral imaging to capture biochemical indicators of ripeness. In addition, optimizing model efficiency for edge deployment will be essential to enable large-scale and sustainable implementation in real-world agricultural environments.

References

[1] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, Art. no. 2674, 2018, doi: 10.3390/s18082674.

[2] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem, and A. M. Saleh, "Evaluation of machine learning approaches for precision farming in smart agriculture system: A comprehensive review," *IEEE Access*, vol. 12, pp. 60155–60184, 2024, doi: 10.1109/ACCESS.2024.3390581.

[3] E. M. B. M. Karunathilake, A. T. Le, S. Heo, Y. S. Chung, and S. Mansoor, "The path to smart farming: Innovations and opportunities in precision agriculture," *Agriculture*, vol. 13, no. 8, Art. no. 1593, 2023, doi: 10.3390/agriculture13081593.

[4] G. Abera, A. M. Ibrahim, S. F. Forsido, and C. G. Kuyu, "Assessment on post-harvest losses of tomato (*Lycopersicon esculentem* Mill.) in selected districts of East Shewa Zone of Ethiopia using a commodity system analysis methodology," *Heliyon*, vol. 6, no. 4, Art. no. e03749, 2020, doi: 10.1016/j.heliyon.2020.e03749.

[5] Food and Agriculture Organization of the United Nations, "FAOSTAT: Crops and livestock products," FAOSTAT, Rome, Italy, 2024. [Online]. Available: https://www.fao.org/faostat/en/#data/QCL

[6] J. Famuyini, O. A. Patrick, and A. Sedara, "Effect of maturity stage on quality and shelf life of tomato (*Lycopersicon esculentum* Mill.) using refrigerator storage system," *Eurasian J. Agric. Res.*, vol. 4, no. 1, pp. 23–44, 2020. [Online]. Available: https://dergipark.org.tr/en/pub/ejar/article/766785

[7] V. Thole *et al.*, "Analysis of tomato post-harvest properties: Fruit color, shelf life, and fungal susceptibility," *Curr. Protoc. Plant Biol.*, vol. 5, no. 2, Art. no. e20108, 2020, doi: 10.1002/cppb.20108.

[8] T. Khatun, A. Razzak, M. S. Islam, and M. S. Uddin, "An extensive real-world in field tomato image dataset involving maturity classification and recognition of fresh and defect tomatoes," *Data Brief*, vol. 51, Art. no. 109688, 2023, doi: 10.1016/j.dib.2023.109688.

[9] G. L. Tenorio and W. Caarls, "Automatic visual estimation of tomato cluster maturity in plant rows," *Mach. Vis. Appl.*, vol. 32, no. 4, Art. no. 78, 2021, doi: 10.1007/s00138-021-01202-9.

[10] M. L. Ali and Z. Zhang, "The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection," *Computers*, vol. 13, no. 12, Art. no. 336, 2024, doi: 10.3390/computers13120336.

[11] N. Rane, "YOLO and Faster R-CNN object detection for smart Industry 4.0 and Industry 5.0: Applications, challenges, and opportunities," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4624206.

[12] J.-W. Chen, W.-J. Lin, H.-J. Cheng, C.-L. Hung, C.-Y. Lin, and S.-P. Chen, "A smartphone-based application for scale pest detection using multiple-object detection methods," *Electronics*, vol. 10, no. 4, Art. no. 372, 2021, doi: 10.3390/electronics10040372.

[13] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, Art. no. 677, 2023, doi: 10.3390/machines11070677.

[14] J. Liu, J. Sun, Y. Wang, X. Liu, Y. Zhang, and H. Fu, "Non-destructive detection of fruit quality: Technologies, applications and prospects," *Foods*, vol. 14, no. 12, Art. no. 2137, 2025, doi: 10.3390/foods14122137.

[15] Z. Yang, Y. Li, Q. Han, H. Wang, C. Li, and Z. Wu, "A method for tomato ripeness recognition and detection based on an improved YOLOv8 model," *Horticulturae*, vol. 11, no. 1, Art. no. 15, 2025, doi: 10.3390/horticulturae11010015.

[16] A. B. Kurina, A. E. Solovieva, I. A. Khrapalova, and A. M. Artemyeva, "Biochemical composition of tomato fruits of various colors," *Vavilov J. Genet. Breed.*, vol. 25, no. 5, pp. 514–527, 2021, doi: 10.18699/VJ21.058.

[17] M. A. Abu Bakar, Z. Zulkifli, M. Mohamad, and S. Ahmad, "GLCM-based feature extraction and CNN approach for fruit freshness detection," in *Proc. 5th Int. Conf. Artificial Intelligence and Data Sciences (AiDAS)*, 2024, pp. 1–6, doi: 10.1109/AiDAS63860.2024.10730167.

[18] D. Worasawate, P. Sakunasinha, and S. Chiangga, "Automatic classification of the ripeness stage of mango fruit using a machine learning approach," *AgriEngineering*, vol. 4, no. 1, pp. 32–47, 2022, doi: 10.3390/agriengineering4010003.

[19] J. Rajesh, M. S. Ashraf, L. Kaur, S. Rout, S. K. Nayak, G. Kaur, and D. R. K. Saikanth, "Application of fuzzy logic in smart agriculture to recognise tomato fruit ripeness," *Int. J. Food Nutr. Sci.*, vol. 11, Special Issue 1, pp. 2360–2367, 2022. [Online]. Available: https://ijfans.org/index.php/Journal/article/view/6468

[20] C. Jiang, K. Miao, Z. Hu, F. Gu, and K. Yi, "Image recognition technology in smart agriculture: A review of current applications, challenges and future prospects," *Processes*, vol. 13, no. 5, Art. no. 1402, 2025, doi: 10.3390/pr13051402.

[21] A. Septiarini, A. Sunyoto, H. Hamdani, A. A. Kasim, F.

 EISSN: 2961-872X ●PISSN: 2961-8924●DOI: 10.56741/IISTR.esl.002559

Utaminingrum, and H. R. Hatta, "Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features," *Sci. Hortic.*, vol. 286, Art. no. 110245, 2021, doi: 10.1016/j.scienta.2021.110245.

[22] N. Iqbal, R. Mumtaz, U. Shafi, and S. M. H. Zaidi, "Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms," *PeerJ Comput. Sci.*, vol. 7, Art. no. e536, 2021, doi: 10.7717/peerj-cs.536.

[23] A. R. Zubair and O. A. Alo, "Grey level co-occurrence matrix (GLCM) based second order statistics for image texture analysis," *arXiv preprint arXiv:2403.04038*, 2024. [Online]. Available: https://arxiv.org/abs/2403.04038

[24] P. Sahu, A. P. Singh, A. Chug, and D. Singh, "A systematic literature review of machine learning techniques deployed in agriculture: A case study of banana crop," *IEEE Access*, vol. 10, pp. 87333–87360, 2022, doi: 10.1109/ACCESS.2022.3199926.

[25] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[26] J. E. Gallagher and E. J. Oughton, "Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications and challenges," *IEEE Access*, vol. 13, pp. 7366–7395, 2025, doi: 10.1109/ACCESS.2025.3526458.

[27] Y. Safari, J. Nakatumba-Nabende, R. Nakasi, and R. Nakibuule, "A review on automated detection and assessment of fruit damage using machine learning," *IEEE Access*, vol. 12, pp. 21358–21381, 2024, doi: 10.1109/ACCESS.2024.3362230.

[28] S. Roy, "Fruit ripeness detection using deep learning," 2025.

[29] M. Bouni, B. Hssina, K. Douzi, and S. Douzi, "Synergistic use of handcrafted and deep learning features for tomato leaf disease classification," *Sci. Rep.*, vol. 14, Art. no. 26822, 2024, doi: 10.1038/s41598-024-71225-5.

[30] A. Kamilaris and F. X. Prenafeta-Boldú, "A review of the use of convolutional neural networks in agriculture," *J. Agric. Sci.*, vol. 156, no. 3, pp. 312–322, 2018, doi: 10.1017/S0021859618000436.