p-ISSN 2715-2871
e-ISSN 2714-5247

Deep Learning Approach for Identifying Fresh and Rotten Chili Peppers Using YOLO

Akbar Sujiwa
Department of Physics
UPN "Veteran" Jatim
Surabaya, Indonesia
akbarsujiwa.ft@upnjatim.ac.id

Nailul Hasan
Department of Physics
UPN "Veteran" Jatim
Surabaya, Indonesia
nailul.hasan.fisika@upnjatim.ac.id

Fajar Timur
Department of Physics
UPN "Veteran" Jatim
Surabaya, Indonesia
timur.ft@upnjatim.ac.id

Reffany C. Rizkiarna
Department of Physics
UPN "Veteran" Jatim
Surabaya, Indonesia
reffany.choiru.ft@upnjatim.ac.id

Abstract— **Chili peppers are a vital agricultural commodity, yet post-harvest quality assessment primarily relies on manual inspection, which is subjective, labor-intensive, and prone to inconsistency. This study proposes a deep learning-based computer vision system using the You Only Look Once (YOLO) framework to automate the identification and classification of fresh and rotten chili peppers. The model was trained using a dataset of 400 web-crawled images, annotated and augmented to handle visual diversity. A novel evaluation strategy was employed using synthetic image generation to simulate real-world scenarios, including neatly arranged grids and randomly distributed objects with varying orientations. Experimental results demonstrate that the proposed model effectively localizes and classifies chili peppers with confidence scores ranging from 0.55 to 0.90. Quantitative evaluation results achieved an accuracy of 0.5048, precision of 0.5516, recall of 0.4580, and mAP@0.5 of 0.4662, indicating moderate detection and classification performance under varying visual conditions. The system successfully distinguishes between fresh and rotten categories even under conditions of intra-class variation, such as discoloration and shriveling. These findings validate the robustness of the YOLO-based approach, offering a promising and efficient solution for automated post-harvest quality control and smart agriculture applications.**

Keywords— Deep Learning, YOLO, Chili Pepper, Quality Assessment, Synthetic Image Generation

I. INTRODUCTION

Chili peppers are widely consumed agricultural products and play a significant role in food industries and daily diets, particularly in developing countries. The freshness and quality of chili peppers strongly influence consumer acceptance, market value, and food safety. Quality degradation caused by microbial activity, mechanical damage, or improper storage conditions leads to visible changes such as discoloration, surface defects, and texture deterioration, resulting in considerable post-harvest losses [1]. Therefore, an accurate and reliable method for distinguishing between fresh and rotten chili peppers is essential.

Quality assessment of agricultural products is traditionally performed through manual visual inspection. Although this method is simple to implement, it is highly subjective, labor-intensive, and inconsistent, especially when applied to large-scale sorting and quality control processes [2]. To address these limitations, computer vision–based approaches have been increasingly investigated for automated quality evaluation in agriculture, as they offer objective, repeatable, and efficient inspection capabilities.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the performance of image-based classification tasks. CNN-based methods are capable of automatically extracting discriminative features such as color, texture, and shape from images, making them well suited for food quality and freshness analysis [3][4][5]. Numerous studies have demonstrated the effectiveness of deep learning models in classifying fruits and vegetables based on ripeness levels, defects, and spoilage conditions.

The You Only Look Once (YOLO) framework is a well-known deep learning model originally developed for real-time object detection. YOLO achieves high processing speed and competitive accuracy by performing object localization and classification in a single forward pass of the network [6][7]. Although YOLO is primarily designed for detection tasks, recent studies have shown that it can be adapted for agricultural quality assessment by associating detected objects with quality-related labels, enabling efficient identification and classification of fresh and rotten produce [8].

Several previous studies have applied the YOLO framework for chili pepper detection and quality assessment; however, most of these works were conducted on a relatively small experimental scale. The chili peppers evaluated were generally limited in number, and the experiments often focused on single chili peppers or small groups of objects per image, with limited quality classes such as ripeness or freshness levels [9], [10] [11]. Although these studies demonstrated the feasibility of YOLO-based approaches for chili pepper analysis, the testing scenarios were still relatively simple and did not fully represent real post-harvest conditions.

The main research gap identified in previous studies is the limited evaluation of YOLO models under complex multi-object conditions, where multiple chili peppers with varying orientations, overlaps, occlusions, and quality conditions appear simultaneously in a single image. In addition, previous works generally relied on conventional datasets and lacked more diverse evaluation strategies capable of testing the robustness and generalization ability

BEST
Journal of Applied Electrical & Science Technology – University of PGRI Adi Buana Surabaya
p-ISSN 2715-2871
e-ISSN 2714-5247

of the trained models in realistic agricultural environments.

To address this gap, this study proposes a YOLO-based deep learning system combined with synthetic image generation for evaluation under more realistic and complex scenarios. The novelty of this research lies in the use of synthetic test images containing multiple fresh and rotten chili peppers with varying object arrangements, orientations, and object combinations to simulate post-harvest conditions more effectively. Furthermore, the proposed model is quantitatively evaluated using accuracy, precision, recall, and mAP@0.5 metrics to provide a comprehensive assessment of detection performance. This approach is expected to improve the robustness of chili pepper quality identification systems for automated post-harvest handling and smart agriculture applications.

II. Method

This section describes the methodology employed in this study to detect chili peppers and classify their quality into fresh and rotten categories using a YOLO-based deep learning approach. The proposed method consists of dataset preparation using web-crawled images, synthetic image generation, model architecture and training, and performance evaluation.

A. Dataset Preparation

The dataset used in this study was collected using a web-crawling approach based on the Bing image search engine. A total of 400 chili pepper images were obtained by querying relevant keywords related to fresh and rotten chili peppers. This crawling strategy enables the collection of diverse image samples with variations in lighting conditions, backgrounds, camera angles, and chili pepper appearances, reflecting real-world visual diversity.

The collected images were manually filtered to remove irrelevant, low-quality, or duplicate samples. Each image was then manually annotated using bounding boxes and labeled into two quality classes: fresh and rotten. The annotation process was conducted using a labeling tool (Figure 1) compatible with the YOLO format to ensure accurate object localization and class assignment.

After annotation, the dataset was resized to a uniform input resolution and converted into YOLO-compatible annotation files. The dataset was subsequently divided into training, validation, and testing subsets according to a predefined ratio to ensure fair and unbiased model evaluation.

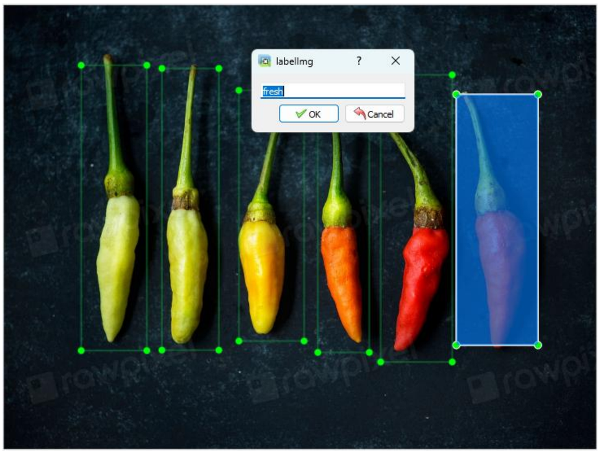

Figure 1. Sample of fresh chilli image labeling process

B. Synthetic Image Generation

To evaluate the robustness of the proposed model under more realistic and complex conditions, this study employs Synthetic Image Generation for the testing phase. Synthetic test images were generated by combining multiple annotated chili pepper objects extracted from the original dataset into a single image. This process produced images containing varying numbers of chili peppers with different spatial arrangements, orientations, overlaps, and combinations of fresh and rotten classes, as illustrated in Figure 2.

The synthetic image generation pipeline consists of several sequential stages. First, chili pepper objects were manually annotated using bounding box labels to separate fresh and rotten chili classes. Each annotated object was then cropped from the original image together with its corresponding object region. Second, augmentation techniques including rotation, scaling, horizontal flipping, and brightness adjustment were applied to each cropped object to increase visual diversity and reduce overfitting. Third, the augmented chili pepper objects were randomly positioned onto a background image to generate a new synthetic scene. During this process, the number of objects, object spacing, orientation, and overlap conditions were randomized to simulate realistic post-harvest environments. The corresponding bounding box coordinates and class labels were automatically recalculated and updated for every generated image. Finally, the resulting synthetic images together with their annotation files were used as testing data for evaluating the YOLO model performance.

The use of synthetic images allows the testing process to simulate real post-harvest scenarios, where multiple chili peppers with diverse quality conditions appear simultaneously in one scene. By separating the training data (original images) from the testing data (synthetic images), the evaluation can better reflect the generalization capability of the trained YOLO model. Furthermore, this approach enables the generation of large-scale and diverse testing data without requiring extensive manual image acquisition and annotation processes.

Figure 2. Examples of Synthetic Test Images Generated by Combining Fresh and Rotten Chili Peppers in a Single Scene.

C. YOLO-Based Model Architecture

The proposed system utilizes the YOLO (You Only Look Once) framework for chili pepper detection and quality classification. YOLO performs object detection and classification in a single forward pass, enabling efficient and real-time inference.

In this study, the YOLO model was configured to detect chili peppers and classify each detected object into one of two categories: fresh or rotten. The network architecture consists of a backbone for feature extraction, a neck for multi-scale feature fusion, and a detection head for bounding box regression and class prediction.

D. Model Training Configuration

The proposed object detection system was developed using Ultralytics YOLO version YOLOv8m as the pretrained model. The training process was conducted using the prepared dataset with several predefined hyperparameters. The input image size was set to 640 × 640 pixels, batch size was configured to 16, and the model was trained for 100 epochs. The AdamW optimizer was employed with an initial learning rate of 0.001 to improve convergence stability during training. In addition, an early stopping mechanism with a patience value of 20 epochs was applied to reduce overfitting and prevent unnecessary training iterations.

Data augmentation techniques such as rotation, scaling, horizontal flipping, translation, and brightness adjustment were applied during training to improve model generalization and robustness under varying visual conditions. These augmentation methods help the model recognize chili peppers with different orientations, sizes, and illumination levels.

The training process optimizes the YOLO loss function, which consists of localization loss, objectness loss, and classification loss. Model performance was continuously monitored using the validation dataset, and the best-performing model was automatically selected for final evaluation based on the validation results.

E. Qualitative Evaluation Based on Bounding Box Scanning

A qualitative evaluation was conducted to assess the ability of the proposed YOLO-based model to detect and classify chili peppers into fresh and rotten categories visually.

In this study, the evaluation was performed using synthetic images under four different scenarios to observe the model's performance in varying conditions:

1. Image containing only fresh chili peppers (12 objects)

Used to evaluate the model's ability to detect all fresh chili peppers and ensure that bounding boxes accurately cover each object.

2. Image containing only rotten chili peppers (12 objects)

Used to assess the model's capability in detecting rotten chili peppers and correctly distinguishing them from fresh chili, with bounding boxes aligned to visual object boundaries.

3. Mixed fresh and rotten chili peppers arranged neatly

Used to evaluate the model's performance in detecting and classifying multiple objects in a structured layout with minimal overlap, facilitating visual inspection.

4. Mixed fresh and rotten chili peppers arranged randomly

Used to assess the model's robustness under randomly positioned objects, while ensuring that all chili peppers remain non-overlapping. This setup simulates realistic post-harvest conditions where chilies are scattered in irregular positions but each object remains clearly distinguishable for detection and classification.

For each variation, the evaluation procedure involved visually inspecting whether the bounding boxes correctly encompassed each chili pepper and whether the assigned class labels corresponded to the observed quality of the objects. This qualitative approach provides insight into the effectiveness of the model in detecting and classifying chili peppers under different visual arrangements.

III. RESULT AND DISCUSSION

This section presents the outcomes of the proposed YOLO-based chili pepper detection and classification system. The results are evaluated using synthetic test images generated from annotated chili peppers, focusing on the qualitative assessment of bounding box placement and class labeling. The discussion highlights the model's performance under different object arrangements and visual conditions, and compares the observations with expectations based on the designed methodology.

A. Detection of 12 Fresh Chili Peppers

Figure 3 illustrates the detection results for 12 fresh chili peppers using the proposed YOLO-based model. Each chili pepper is enclosed within a bounding box, with class labels and confidence scores displayed above each object.

The model successfully detected all 12 chili peppers, and each bounding box accurately covered the corresponding object without overlapping adjacent chili peppers. Confidence scores ranged from 0.56 to 0.90, indicating reliable class prediction for all objects. Visual inspection confirms that the bounding boxes follow the shape and orientation of each chili pepper closely, demonstrating the model's effectiveness in detecting multiple fresh chili peppers in a single frame.

In addition to qualitative analysis, quantitative evaluation metrics were also employed to measure the detection performance of the proposed model. The evaluation results produced an accuracy value of 0.5048, precision of 0.5516, recall of 0.4580, and mAP@0.5 of 0.4662. These results indicate that the model was capable of detecting and classifying chili peppers with moderate performance under the tested conditions. The precision value shows that more than half of the detected objects were correctly classified, while the recall value indicates that several chili peppers were still missed during detection. Furthermore, the obtained mAP@0.5 value demonstrates that the model was able to localize objects reasonably well under varying object positions and orientations. The confusion matrix analysis also revealed that most objects were correctly identified, with some false positives and false negatives occurring in visually similar object conditions.

These results show that the proposed system can handle images containing multiple objects of the same class while maintaining acceptable localization and labeling performance, which is essential for practical applications in post-harvest quality assessment and automated sorting systems.

B. Detection of 12 Rotten Chili Peppers

Figure 4 demonstrates the detection results for 12 rotten chili peppers using the developed YOLO-based model. Similar to the previous experiment, the system identified and localized the objects within the frame, assigning the class label "rotten" to each detected chili pepper.

The model successfully detected all 12 rotten chili peppers, achieving a 100% detection rate for this sample set. The bounding boxes accurately encapsulate the varying shapes and sizes of the decayed peppers. Confidence scores for the detection ranged from 0.55 to 0.86. While these scores are slightly more varied compared to the fresh chili peppers, they remain above the detection threshold, confirming the model's capability to distinguish spoiled produce.

It is important to note that the visual characteristics of the rotten peppers in Figure 3 vary significantly, ranging from slight discoloration to severe shriveling and textural changes. Despite this intra-class variance, the model maintained consistent classification. This suggests that the model has learned robust features associated with spoilage, which is critical for automating the separation of defective items in agricultural sorting lines.

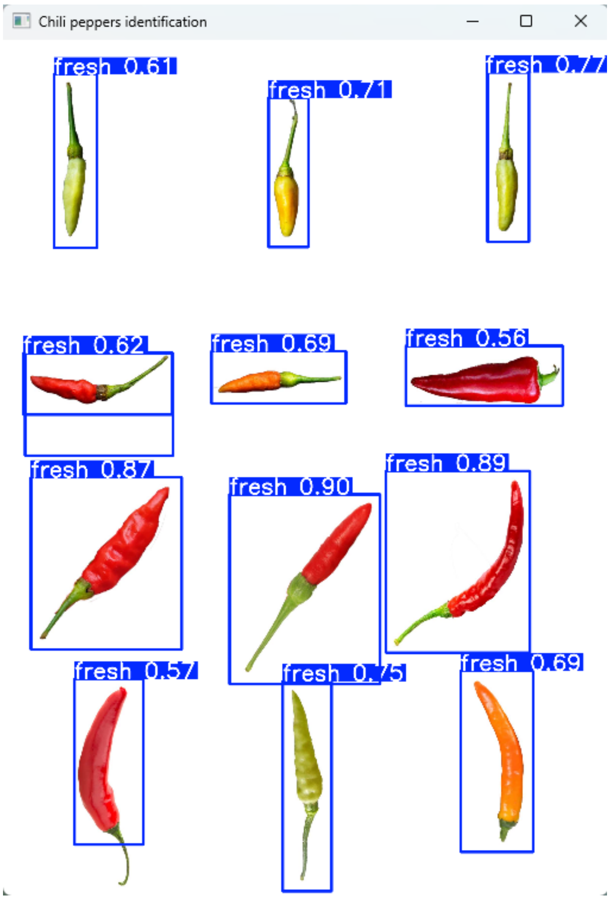

Figure 3. Image of 12 fresh chili peppers with detection bounding boxes

C. Detection of Mixed Fresh and Rotten Chili Peppers

To evaluate the model's discriminative capabilities in a heterogeneous environment, a third experiment was conducted using a mixed dataset. Figure 5 presents the detection results for a collection of 12 chili peppers consisting of both fresh and rotten samples arranged in a grid.

The proposed YOLO-based model successfully localized and classified all 12 objects within the frame. As shown in the figure, the system distinctly separated the classes, assigning blue bounding boxes to "fresh" chili peppers and cyan bounding boxes to "rotten" ones. In this specific test case, the model correctly identified 7 fresh peppers and 5 rotten peppers.

The confidence scores varied across the samples. For the "rotten" class, scores ranged from 0.63 to 0.83, showing consistent detection even when placed adjacent to fresh samples. For the "fresh" class, scores generally ranged between 0.57 and 0.86, with one outlier detected at a lower confidence of 0.26. Despite this lower confidence score on one instance, the class prediction remained correct. This successful differentiation in a mixed-object scenario demonstrates the model's robustness and its potential for deployment in real-world sorting systems where fresh and defective produce are processed simultaneously.

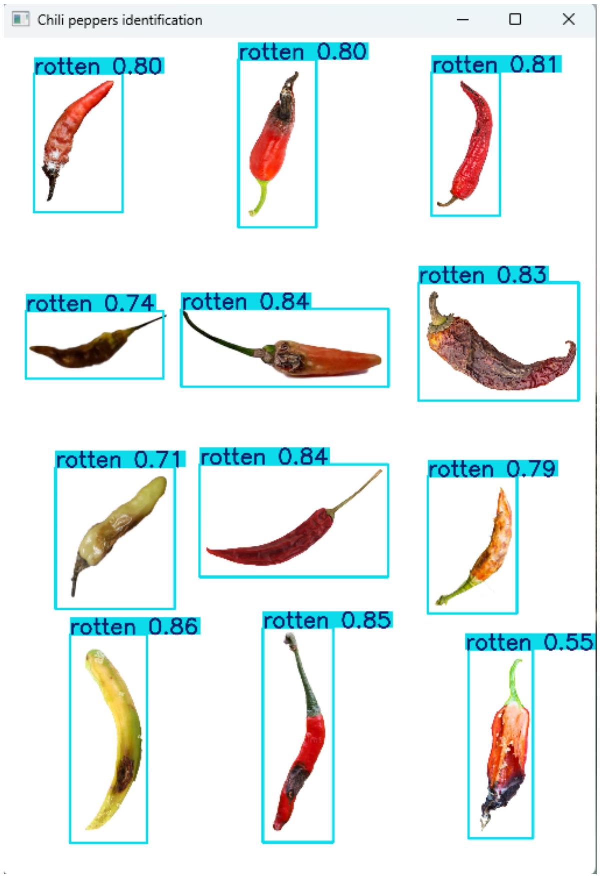

Figure 4. Image of 12 rotten chili peppers with detection bounding boxes

D. Detection of Randomly Arranged Fresh and Rotten Chili Peppers

To assess the model's performance in a more realistic and complex environment, a final experiment was conducted using a randomly arranged pile of chili peppers. Figure 6 displays the detection results for mixed fresh and rotten chili peppers positioned with random orientations and significant overlap.

As observed in the figure, the model demonstrated a capability to detect and classify objects even under conditions of occlusion and clutter. The system identified multiple overlapping instances, distinguishing between "fresh" (blue bounding boxes) and "rotten" (cyan bounding boxes) peppers within the dense arrangement.

The confidence scores in this scenario showed higher variance compared to the neatly arranged experiments. While distinct objects were detected with high confidence (e.g., 0.82 for a fresh pepper and 0.85 for a rotten pepper), partially occluded or smaller objects yielded lower confidence scores, such as 0.29 for a fresh sample and 0.26 for a rotten sample.

Despite the visual complexity, the model successfully applied Non-Maximum Suppression (NMS) to differentiate between individual peppers that were touching or stacked. This result highlights the model's potential for real-world applications where chili peppers are often processed in bulk or piles, although it also indicates that heavy occlusion remains a challenging factor that influences the model's prediction confidence.

IV. Conclusion

This study successfully implemented and evaluated a YOLO-based computer vision system for the detection and quality classification of chili peppers. The experimental results demonstrate the model's efficacy in distinguishing between "fresh" and "rotten" classes under various arrangements.

The system exhibited high precision and confidence in detecting single-class arrangements, accurately identifying both fresh and rotten samples with consistent bounding box placement. Notably, the model demonstrated robustness in handling intra-class variations, effectively recognizing different visual manifestations of spoilage, such as discoloration and shriveling.

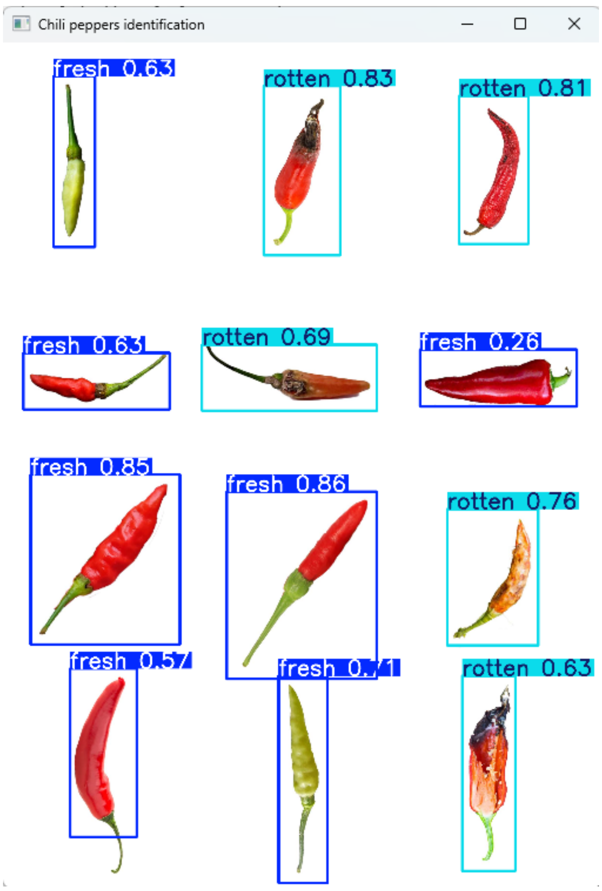

Figure 5. Detection results on a mixed set of fresh and rotten chili peppers

In mixed-class scenarios, the model maintained its discriminative capability, correctly classifying adjacent fresh and rotten objects. Furthermore, the final experiment on randomly arranged piles highlighted the model's potential for real-world application; despite the challenges posed by occlusion and clutter, the system successfully utilized Non-Maximum Suppression to detect individual peppers within a dense arrangement. Although confidence scores varied in highly occluded instances, the detection remained accurate.

Overall, the proposed system shows significant promise for deployment in automated post-harvest processes. It offers a reliable solution for sorting agricultural produce, potentially reducing manual labor and enhancing the efficiency of quality control systems.

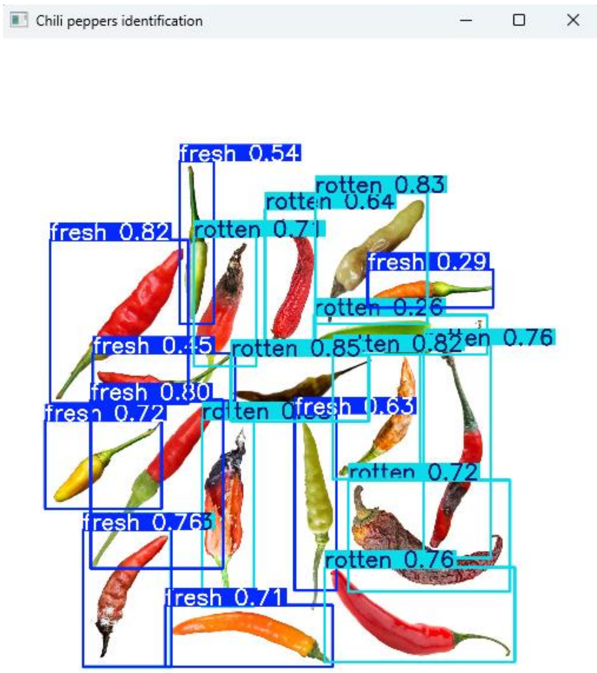

Figure 6. Detection results on a randomly arranged and cluttered pile of chili peppers

REFERENCES

[1] F. and A. O. of the U. Nations, “The State of Food and Agriculture 2019: Moving Forward on Food Loss and Waste Reduction,” Oct. 2019, doi: 10.18356/32F21F8C-EN.

[2] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, “Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables,” *Food Bioproc. Tech.*, vol. 4, no. 4, pp. 487–504, May 2011, doi: 10.1007/S11947-010-0411-8.

[3] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/NATURE14539.

[4] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Comput. Electron. Agric.*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/J.COMPAG.2018.02.016.

[5] A. Sujiwa and H. P. Prasetyo, “MEASUREMENTS ANALYSIS OF MQ-6 LPG GAS LEAK DETECTION DEVICES FOR DISTANCE FACTORS,” *Best : Journal of Applied Electrical, Science and Technology*, vol. 2, no. 1, pp. 6–9, Mar. 2020, doi: 10.36456/best.vol2.no1.2578.

[6] J. S. D. R. G. A. F. Redmon, “(YOLO) You Only Look Once,” *Cvpr*, vol. 2016-December, pp. 779–788, Dec. 2016, doi: 10.1109/CVPR.2016.91.

[7] A. Sujiwa, N. Hasan, and N. M. Aruan, “TensorFlow Based AI Training for Translating Indonesian Sign Language (BISINDO),” *Best : Journal of Applied Electrical, Science and Technology*, vol. 7, no. 1, pp. 29–34, Mar. 2025, doi: 10.36456/BEST.VOL7.NO1.10312.

[8] F. Xiao, H. Wang, Y. Xu, and R. Zhang, “Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review,” *Agronomy 2023, Vol. 13, Page 1625*, vol. 13, no. 6, p. 1625, Jun. 2023, doi: 10.3390/AGRONOMY13061625.

[9] N. Y. D. Rahmawati and F. P. Rochim, “Chili Ripeness Level Detection with YOLOv8 and Fuzzy Logic for Harvest Decision Making,” *Jurnal Teknik Pertanian Lampung (Journal of Agricultural Engineering)*, vol. 14, no. 5, pp. 1807–1818, Oct. 2025, doi: 10.23960/JTEPL.V14I5.1807-1818.

[10] N. Ma, Y. Wu, Y. Bo, and H. Yan, “Chili Pepper Object Detection Method Based on Improved YOLOv8n,” *Plants 2024, Vol. 13, Page 2402*, vol. 13, no. 17, p. 2402, Aug. 2024, doi: 10.3390/PLANTS13172402.

[11] A. Sujiwa and R. R. Dianto, “INFUSION MONITORING SYSTEM FOR PATIENTS BASED ON THE INTERNET OF THINGS (IOT) WITH ANDROID NOTIFICATION SYSTEM,” *Best : Journal of Applied Electrical, Science and Technology*, vol. 4, no. 2, pp. 41–46, Nov. 2022, doi: 10.36456/best.vol4.no2.6172.