BEST
Journal of Applied Electrical & Science Technology – University of PGRI Adi Buana Surabaya
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*—In Indonesia, chili peppers are a vital agricultural commodity, widely used in various local dishes. Chili peppers are widely sold in modern markets and traditional markets. In some traditional markets, sellers sometimes do not store them in suitable conditions, thereby affecting the quality of the chili peppers. Sometimes, buyers are unable to select the specific chilies being sold wholesale, so they often end up with chilies of inferior quality. Therefore, this study proposes a deep learning-based computer vision system using the You Only Look Once (YOLO) framework to distinguish between fresh and rotten chili peppers. The model that we use was trained using 400 datasets from web-crawled images. To simulate real-world scenarios, during testing, some synthetic images were arranged in both regular and random patterns, as well as various chili 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 result of the experiment shows that the system success for distinguishes between fresh and rotten chilli pepper in some conditions of image variation, such as the angular orientation of the chilli peppers and their proximity. This shows that the approach using YOLO provides very promising results for application in the agricultural market.**

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

I. INTRODUCTION

Chili peppers are very popular and widely consumed in Indonesia and play a significant role in food industries and daily consumption, particularly in developing countries. The freshness of a chili pepper significantly affects the dish, market value, and food safety. Its quality is also determined by microbiological factors, mechanical damage, or poor storage conditions leads to visible changes such as chili color change, skin defect, and texture deterioration, resulting in considerable decline in sales [1]. Therefore, an accurate and reliable method for distinguishing between fresh and rotten chili peppers is very important.

Quality assessment of agricultural products is traditionally performed through manual visual inspection. This method is simple to implement, but has weakness such as highly subjective, labor-intensive, and inconsistent when applied to large-scale sorting and quality control processes [2]. To address these limitations, many researchers have studied and developed automated computer vision technology to evaluate quality in agriculture, as they offer objective, repeatable, and efficient inspection capabilities.

The current development of Convolutional Neural Networks (CNNs) as a deep learning method has significantly improved the performance of image-based classification tasks. CNN-based methods are capable of automatically running differentiating feature such as color, texture, and shape from images, making them very suitable for food quality and freshness analysis [3][4][5]. Many studies have shown 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) is a framework long known as a learning model for object detection. YOLO offers advantages in data processing speed as well as high accuracy in object localization and classification in a single forward pass of the network [6][7]. Although the primary YOLO design is used as a detection tool, recent studies have shown that it can be used in the agricultural sector to determine the quality of fruits and vegetables, enables the efficient identification and classification of fresh and spoiled agricultural products [8].

YOLO framework has been applied to several previous studies have especially for chili pepper for detection and quality determination; however, most of these works were conducted on a relatively small experimental scale. The number of chili pepper images observed was not very large, and more often simply observe a single chili pepper in various forms, with limited quality classes such as ripeness or freshness levels [9], [10] [11]. Although numerous previous studies have addressed chili freshness using YOLO, the testing scenario is also quite simple and do not adequately reflect actual post-harvest conditions.

The main research gap between this study and other previous studies is the limited evaluation of YOLO models, such as the positioning of the chilies across various rotations, the arrangement of chilies within the test image layout, and the condition—fresh or rotten—of the chilies within a single test image. In addition, previous works generally used conventional datasets and limited

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

variation in test data, which influences the testing of the robustness and generalization ability of the trained models in realistic agricultural environments.

To address this problem, in this research proposes a deep learning system using YOLO combined with synthetic image generation for evaluation under more realistic and complex scenarios. The novelty of this research lies in the use of synthetic images—composed using various positions, sizes, and orientations of fresh and rotten chili peppers—to simulate piles of chili peppers in a market. Furthermore, the proposed model is quantitatively evaluated using accuracy, precision, recall, and mAP@0.5 metrics to provide a comprehensive assessment of detection performance. With this approach it is hoped that it can improve identification ability of chili pepper quality in smart agriculture applications.

II. METHOD

This methodology section outlines the steps for using a YOLO-based deep learning approach to determine the quality of chili peppers (fresh versus rotten). The proposed method involves dataset preparation using web-crawled images, the generation of synthetic images resembling piles of chili peppers, model training, and performance testing of the trained model.

A. Dataset Preparation

The dataset in this research was created by downloading images using web-crawling machine like Bing image search engine, it will automatically searches from the internet for images of chili peppers. The dataset that consisting of 400 images of chilies—both fresh and rotten—was selected. This technique yields a variety of chili pepper images across diverse environmental conditions—such as lighting, backgrounds, shapes, sizes, quantities, clustering, and chili varieties—designed to approximate real-world scenarios.

The collected images were manually filtered to remove irrelevant, low-quality, or duplicate samples. Naturally, not all image data meets expectations in terms of quality; filtering is required to remove irrelevant images, duplicates, cartoons, and images unrelated to chili peppers. Once 400 images were successfully obtained, object labeling was performed using labeling tools (Figure 1) compatible with the YOLO format to ensure the accuracy of object positioning and classification.

Since the images in the database are not of uniform size, resolution adjustments are required. The images from the existing dataset are sorted and distributed into separate folders to distinguish between the data intended for training, validation, and testing.



Figure 1. Sample of fresh chilli image labeling process

B. Synthetic Image Generation

To assess the reliability of the trained machine learning model, this study utilizes synthetic images as test data. This synthetic image was created by combining several images from a dataset into a single image. The variations in the chili images provided include color, size, type, position, and rotation. With this variation we will get image that contains varying numbers of chili peppers with some arrangements, such as orientations, color, and combinations of fresh and rotten classes, as illustrated in Figure 2.

Next steps for creating synthetic images are we are creating several variations. First, for the first test image, we created a composition featuring a total of nine fresh chili peppers, arranged with uniform spacing and variations in color, size, and orientation. Second, as with the first step, a second test image was created using the same variations, but featuring a rotten chili; this was intended to determine whether the trained model could distinguish between fresh and rotten chilies. Third, for this third test image, we are trying a combination of fresh and rotten chilies arranged within a single test image. Fourth, in this image, the quantities of rotten and fresh chilies are combined; however, unlike the first image, the chilies here are arranged unevenly, similar to Figure 2. The expectation is that, during testing, the system will output a bounding box class that matches the actual condition of the chili peppers, even when they are in a random pile or arrangement.

This approach is expected to simulate the actual conditions of chili peppers sold on the market, so that the resulting image analysis data can be applied in real-world scenarios, such as when the peppers are piled up or stored in containers. With this separation (the training data (original images) from the testing data (synthetic images)), aims to determine whether the model is capable of detecting images or conditions outside of its training dataset. Furthermore, This will significantly optimize time and resources, as we do not need to set up real-world conditions.

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



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 system being developed utilizes the YOLO (You Only Look Once) framework from Ultralytics. This framework will be used to create a machine learning model that distinguishes between fresh and rotten chili peppers. The advantage of YOLO is its ability to perform object detection and classification in a single forward pass, making it highly efficient and capable of real-time operation.

In general, the YOLO model in this study was modified and configured solely to detect the condition of the chili peppers (fresh or rotten), without classifying their size or type. The main network architecture consists of a backbone that use for feature extraction, a neck for multi-scale feature fusion, and a detection head for showing bounding box regression and class prediction.

D. Model Training Configuration

An object detection system developed using Ultralytics YOLOv8m as the pre-trained model version. Although there are various versions of YOLO, we are using this one because it is fast and accurate, given the available hardware. The training process was conducted using the prepared dataset with several predefined hyperparameters. Before being used for training, all images in the dataset must be resized to 640 × 640 pixels so that the image dimensions remain constant during training. The batch size was configured to 16, and the model was trained for 100 epochs. To improve accuracy, the AdamW optimizer is also used to prevent overfitting, with an initial learning rate of 0.001 to improve convergence stability during training. Additionally, to prevent overfitting caused by an excessive number of epochs, the maximum number of epochs is limited to 20.

Data augmentation techniques are also employed to enhance model training quality by automatically applying transformations via code; these transformations include rotation, scaling, horizontal flipping, translation, and brightness adjustment. These augmentation methods help the model recognize chili peppers with different orientations, sizes, and illumination levels.

Retraining the YOLO model will improve the model's accuracy and also enable it to recognize classes that were not present in the base model. 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 based on visual observation to determine the success of the trained YOLO model, assessed by the appearance of classes and bounding boxes that correspond to the actual conditions.

In this study, the evaluation scenario utilizes four previously generated synthetic images to assess the performance of the developed model:

1. Image consisting of 12 fresh chili peppers

This image is used to evaluate whether the trained model is capable of detecting variations in color, size, type, and simple orientation of fresh chili peppers.

2. Image consisting of 12 rotten chili peppers

This image is used to evaluate whether the trained model is capable of detecting variations in color, size, type, and simple orientation of rotten chili peppers.

3. Mixed fresh and rotten chili peppers arranged neatly

This image is used to evaluate the model's ability to distinguish between fresh and rotten chilies arranged in an orderly manner within a single image.

4. Mixed fresh and rotten chili peppers arranged randomly

This image is used to simulate the actual condition of a chili pile consisting of a mixture of fresh and rotten chilies, with their positions and orientations arranged randomly.

For each evaluation, a visual check is performed by observing the appearance of the bounding box relative to the image, as well as the correspondence between the class and the image of the chili pepper being tested. Based on these results, the model's behavior in detecting chili objects can be analyzed in relation to the introduced variations.

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

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

The results of testing using a synthetic image containing 12 fresh chili peppers are shown in Figure 3. From the figure, it can be observed that the detected chili peppers are enclosed within bounding boxes, and their condition is indicated by the class name displayed above, along with the corresponding confidence score.

It can be seen here that the trained model is capable of detecting the presence of chili objects, indicated by bounding boxes. The model is also capable of accurately defining the chili class, as evidenced by all chilies being identified as fresh chilies. Confidence scores ranging from 0.56 to 0.90 indicate that the model is effective at confirming the detected object is a fresh chili pepper. This also indicates that the model is capable of effective detection, despite variations in object orientation and the large number of chili peppers being analyzed.

To reinforce the qualitative analysis, a quantitative evaluation was also conducted by performing a software-based model performance test. The results of this quantitative testing were obtained accuracy value of 0.5048, precision of 0.5516, recall of 0.4580, and mAP@0.5 of 0.4662. Based on the results of the conducted testing, the performance of the trained model falls within the moderate range. The precision value indicates that the model is capable of correctly detecting more than half of the tested objects. Meanwhile, the recall value indicates that the model still occasionally fails to identify chili objects. Meanwhile, the mAP@0.5 value indicates that the model is capable of localizing object positions, even across various positions and orientations. Furthermore, the confusion matrix analysis indicates that the developed model is capable of detecting the majority of chili objects, although there are still some false positives and false negatives involving objects that appear similar.

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*

The results of the rotten chili detection test are shown in Figure 4. This image features 12 images of rotten chilies arranged neatly, with variations in color and orientation. As with the previous test results, the successfully identified chili peppers are marked with a bounding box, the object class name, and a confidence score at the top.

From the image, it can be observed that the trained model is capable of recognizing the entire rotten chili object. The bundling box also precisely encloses the chili area without missing any part. The confidence score ranges from 0.55 to 0.86. When compared to the confidence score for fresh chili peppers, the values do not differ significantly. Thus, it can be said that the model's performance in detecting fresh and rotten chilies is relatively similar.

Based on the analysis results from these two tests, it can be confirmed that the developed model possesses a satisfactory ability to distinguish objects according to the trained class parameters. So, it is predicted that it can be applied to more complex testing in the future.

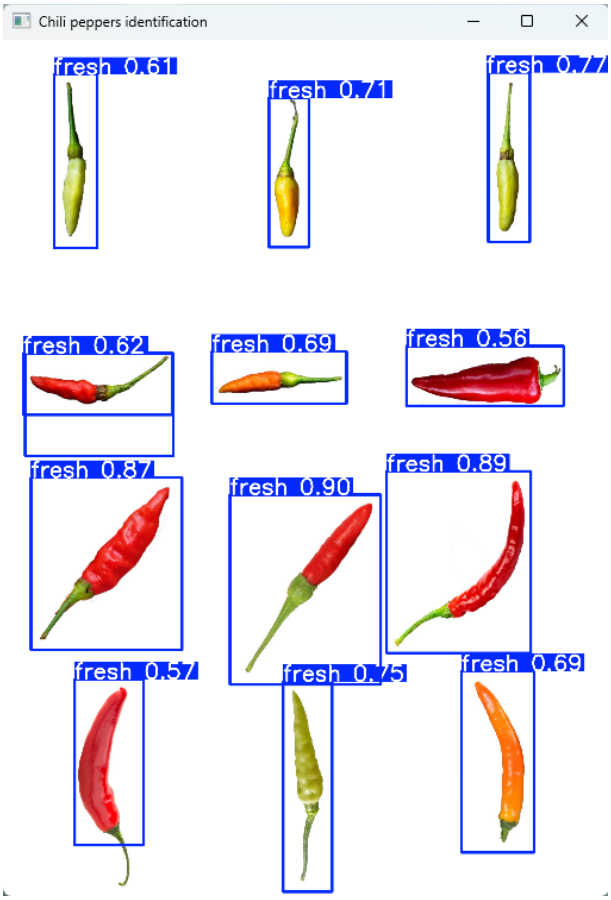


Figure 3. Image of 12 fresh chili peppers with detection bounding boxes

C. *Detection of Mixed Fresh and Rotten Chili Peppers*

While the two previous tests involved only homogeneous images (consisting entirely of either fresh or rotten chilies), this third test involved a mixture of chili types. This is done by replacing some of the fresh chili peppers with rotten ones. The goal is to ensure that the trained model is truly capable of distinguishing between the chili classes when they are combined. However, in this test, the chili peppers were still neatly arranged to reduce the model's analysis load.

Figure 5 shows the results of this test. It can be seen that the model is capable of localization by placing bounding boxes around all the objects present. All images have been classified, and their confidence scores can be displayed. Based on this, the number of fresh chilies can be calculated as 7, while the number of rotten chilies is 5. Thus, it can be said that the model can accurately distinguish between chili classes without error.

Meanwhile, the confidence scores themselves range from 0.26 to 0.83. Despite the relatively small values, the

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

model is still able to correctly define the object classes. Upon closer examination of the previous testing, the objects that yielded low values in this test were also the ones that yielded low values in the earlier test. This indicates that the model has not been well-trained, which could be due to a lack of test images or images sharing similar characteristics across both classes. Furthermore, the model has demonstrated an initial ability to distinguish between different classes under mixed conditions, making it ready for testing under even more complex conditions.

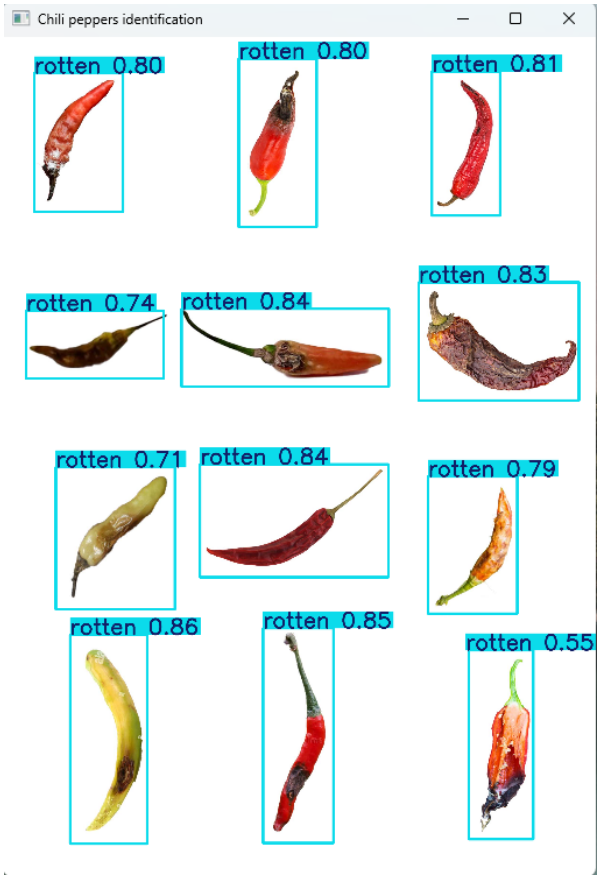


Figure 4. Image of 12 rotten chili peppers with detection bounding boxes

D. Detection of Randomly Arranged Fresh and Rotten Chili Peppers

In this test, the difficulty level of object detection was increased by randomly and asymmetrically positioning fresh and rotten chilies. This is done to test the model's performance under conditions that resemble real-world scenarios. with the proviso that the position of the chilies is maintained so that none overlap or cover one another. The results of this test can be seen in Figure 6.

Based on the image, it can be analyzed that although the bounding box identification results appear messy and overlapping, all objects are enclosed within bounding boxes. All images already have their respective classes and confidence scores assigned, with none left out.

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.

However, upon closer inspection, there is one chili pepper that is incorrectly classified. Regarding the chili pepper in the center of the image with a score of 0.26—which should have been classified as fresh—the model incorrectly classified it as rotten. Thus, based on this test, calculating the classification accuracy percentage relative to the total number of objects yields an accuracy of 94.74%.

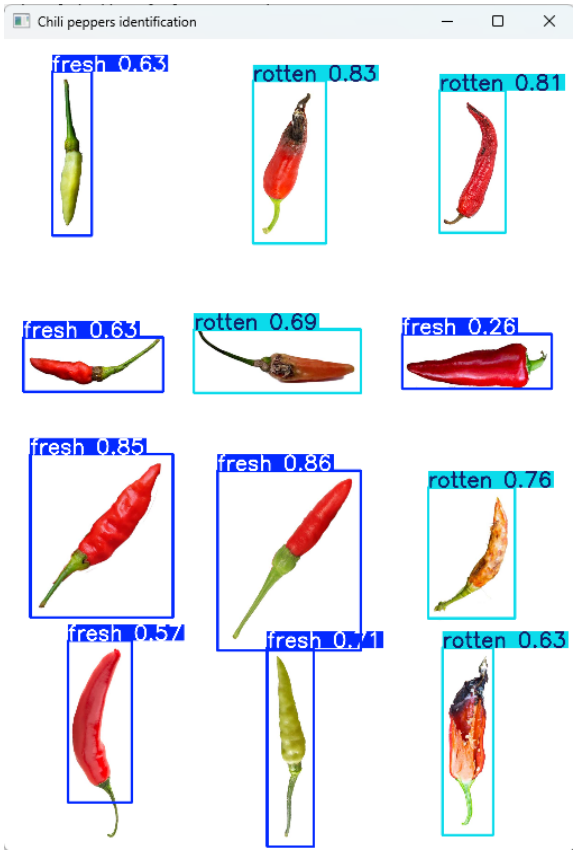


Figure 5. Detection results on a mixed set of fresh and rotten chili peppers

IV. CONCLUSION

This research successfully achieved automated identification—specifically of chili peppers—using a YOLO-based machine learning approach to determine their quality, distinguishing between fresh and rotten states. Test results also indicate that the trained YOLO model is capable of distinguishing between fresh and rotten chili peppers across various chili orientations.

The model analysis results also demonstrate a high level of reading precision; this is quantitatively evidenced by peak bounding box confidence scores of 0.82 for fresh chilies and 0.85 for rotten chilies. The model also demonstrates reliability in distinguishing chili peppers

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

across variations in color, size, type, and orientation, without making a single error in classifying the detected chili peppers.

In the mixed testing scenario, qualitative measurements of the model demonstrate good performance. Because there was one error in classifying the chili peppers as rotten when they should have been classified as fresh. The confidence score for the misidentification result is 0.26, indicating that the model struggles significantly to define the object being analyzed.

Overall, the developed system demonstrates good accuracy and consistency in distinguishing chili quality. Based on these results, it can be concluded that this system is suitable for agricultural automation, particularly for classifying chili pepper quality. Consequently, it has the potential to increase production and reduce reliance on manual sorting, which lacks reliability and efficiency.

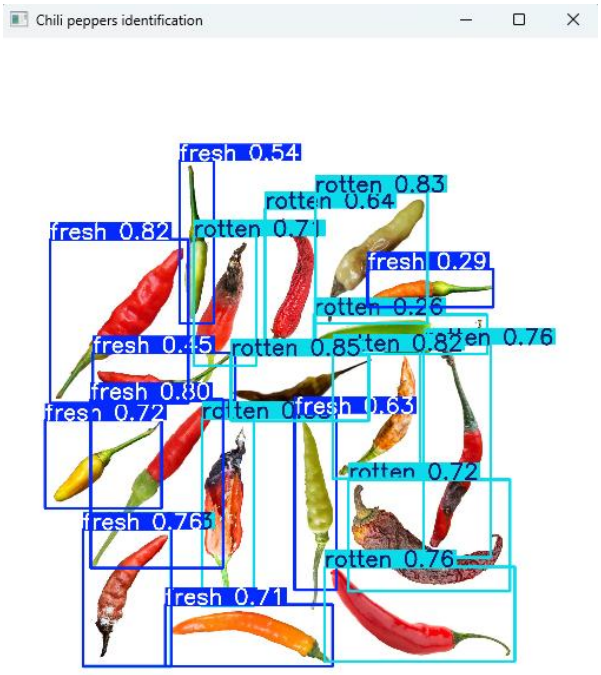


Figure 6. Detection results on a randomly arranged and cluttered pile of chili peppers

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.