2023 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) | 979-8-3503-4389-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRAMET60171.2023.10366571

Ripeness Classification of Cavendish Bananas using Multi-object Detection Approach

Dito Eka Cahya
Research Centre for Electronics
National Research and
Innovation Agency (BRIN)
Bandung, Indonesia
dito001@brin.go.id

Zaid Cahya
Research Centre for Electronics
National Research and
Innovation Agency (BRIN)
Bandung, Indonesia
zaid002@brin.go.id

Heru Taufiqurrohman
Research Centre for Electronics
National Research and
Innovation Agency (BRIN)
Bandung, Indonesia
heru022@brin.go.id

Herdiarti Destika
Research Centre for Agro Industry
National Research and
Innovation Agency (BRIN)
Bogor, Indonesia
herd003@brin.go.id

Mohammad Nafila Alfa
Research Centre for Agro Industry
National Research and
Innovation Agency (BRIN)
Bogor, Indonesia
moha060@brin.go.id

Tantry Eko Putri
Research Centre for Agro Industry
National Research and
Innovation Agency (BRIN)
Bogor, Indonesia
tant003@brin.go.id

***Abstract*—Cavendish bananas, recognized for their global economic and nutritional value, often pose challenges in ensuring consistent ripeness quality throughout the supply chain, which is essential for efficient post-harvest management and ensuring optimal consumer experience. Traditional methods, reliant on human visual inspection, are subjective and inconsistent. This paper proposes a novel multi-object detection approach for accurately classifying the ripeness of multiple Cavendish bananas in an image using computer vision and deep learning techniques. We employed the Yolov5 model, a state-of-the-art object detection architecture using a standardized ripeness classification system to simultaneously identify and categorize every banana in an image. To train the model, a diverse dataset of 600 Cavendish banana bunches at different ripeness stages was collected, augmented, and annotated with ground truth labels. Training result shows that the multi-object detection network successfully delineates the bananas in input images while simultaneously predicting their ripeness classes. The proposed approach achieves 98.8% mean average precision, 90.5% precision, and 92.6% recall, even when the test images contain overlapping bananas and diverse background colors. The compact size of the object detection model, made it applicable to an embedded system, enabling real-time ripeness classification of Cavendish bananas with simple image uploads from handheld devices. This research contributes to the advancement of agricultural technology and opens avenues for future studies in fruit ripeness analysis and food quality assessment.**

***Index Terms*—Multi-object detection, Banana ripeness classification, Computer vision, Deep learning techniques, Yolov5 model, Fruit ripeness analysis**

The research is funded by Rumah Program Teknologi Tepat Guna, Agriculture and Food Research Organization, National Agency for Research and Innovation RI, and Rumah Program Artificial Intelligence dan Big Data, Electronics and Informatics Research Organization, National Agency for Research and Innovation RI

I. Introduction & Related Works

Cavendish bananas, the world's most consumed and traded fruit, hold immense economic and nutritional significance according to the FAO [1]. However, ensuring their optimal quality and ripeness throughout the supply chain remains challenging for stakeholders in the agricultural and food industries [2]. Accurate and efficient ripeness classification is crucial to prevent premature or overripe fruit from reaching consumers, reduce food waste, and enhance overall product quality and customer satisfaction [3].

The accurate assessment of fruit ripeness plays a crucial role in the agricultural industry, particularly in the case of Cavendish bananas [4]. The ripeness classification of bananas is essential for growers, distributors, and retailers to ensure optimal harvesting, storage, and distribution practices [5]. Traditionally, this classification has been performed manually, relying on subjective visual inspection by human experts. However, this approach is time-consuming, labor-intensive, and prone to inconsistencies. In recent years, computer vision and deep learning advancements have shown promising results in automating fruit quality evaluation [6]. Object detection algorithms, in particular, have proven effective in recognizing and localizing objects of interest in images, making them suitable candidates for automating the ripeness classification of Cavendish bananas [7].

Several attempts at banana ripeness classification using computer vision and various machine learning methods have been done in the past. For example, Jiangong et al. [8] approached banana classification with transfer learning. They used a GoogleNet pre-trained model and trained it on 618 locally sourced banana images, which resulted in an impressive accuracy of 98%. On the other hand, Mazen and Nashat's [9] research classifies 4 ripeness levels of bananas using a shallow Convolutional Neural Network (CNN) trained on a dataset of 300 Egyptian bananas which achieved 97% accuracy. Saragih et al. [10] then trained MobileNet-V2 and NASNetMobile pre-trained model to classify ripeness levels using the dataset collected by Mazen and Nashat [9]. They experimented with varying hyper-parameters, including the number of epochs and unfrozen layers, achieving an accuracy of 96%. Zhang et al. [11] developed a seven-class banana classifier using three convolutional neural networks (CNNs) to process original, positive, and negative image inputs simultaneously. Although their method has some drawbacks in the real-time environment, it can reach an accuracy of 94% in the testing environment.

Our previous research [12] improves past works by introducing overripen stage to the classification class, a more practical classification model, and a hyper-parameter tuning approach to increase the ripeness classification performance. Nevertheless, the majority of previous research, including our own, has focused on the classification of individual bananas. This approach may have limitations in practical applications, as bananas are typically encountered in bunches, whether in commercial markets or agricultural plantations. Phoophuangpairoj et al.'s recent research [13] is the closest attempt for determining the ripening stage of a banana bunch. They employed a CNN model to categorize a bunch of bananas into five ripeness classes and achieved 89% accuracy. However, it is noteworthy that within a single bunch, each banana may exhibit a different level of ripeness due to factors such as size, position within the bunch, and exposure to environmental conditions. As such, it is essential to assess the ripeness of individual bananas within a bunch to support timely decision-making for storage, distribution, and marketing strategies, which will improve the overall quality and longevity of bananas.

To tackle the aforementioned research gap, this paper presents a novel multi-object detection approach that leverages cutting-edge computer vision techniques and deep neural networks to classify the ripeness of multiple Cavendish bananas individually. By simultaneously identifying and categorizing individual bananas within images, our proposed method aims to provide a reliable and automated solution for ripeness assessment during the storage and distribution of bananas.

II. Methodology

A. Dataset Collection

To gather appropriate training data for the multi-object detection model, a diverse dataset of Cavendish bananas at various ripeness stages was collected for this study. The banana samples were sourced from local markets and then manually inspected and sorted into distinct ripeness classes, including overripe stages. Before conducting this research, we have developed an intelligent fruit storage chamber that can automatically record environment parameters during fruit ripening process, such as temperature, humidity, CO2, O2,

Fig. 1. Ripening testbed for dataset collection

ethylene, fruit mass, and fruit images (Fig. 1. For the purpose of this research, high-resolution camera installed in the storage chamber was used to capture images of four Cavendish bananas periodically during their natural ripening process in a standardized lighting and background. In total, we recorded 600 banana images which were then annotated with bounding boxes and given ripeness stage labels by experienced food researchers in our team as shown in Fig. 2.

B. Data Pre-processing

The dataset that we collected were processed further with data augmentation technique to increase the generalization of the object detection model [14]. The following are the pre-processing and augmentation steps used in this research:

- Resize the image to the model requirement, which is 640x640 pixels
- Slightly modify original dataset with the following filters:
  - Random Horizontal and Vertical Flip
  - Random Rotation between -20° and 20°
  - Random Shear ±21° Horizontal, ±21° Vertical
  - Random Brightness Between -10% and +10%
  - Random Mosaic

Using these image processing techniques, we managed to increase the number of images in the augmented dataset to four times the original dataset which is 2400 images. The augmented image samples can be seen in Fig. 3. The original images in Fig. 2 only contain four bananas placed in a similar arrangement that matures over time. Thus, the shape and position of the bananas in the image are consistent, with only variations in color and texture of the bananas. The augmentation technique enriches the original images by changing the

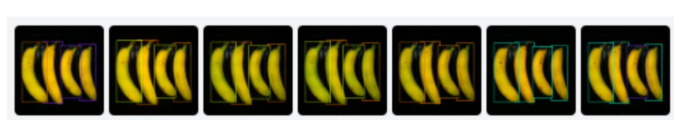


Fig. 2. Data collected from ripening testbed

Fig. 3. Data Augmentation Results

position, angle, brightness, and even creating a mosaic based on the images. Therefore, the augmented dataset was later used for the model training to reduce the model's overfitting, making it effective to classify unseen banana images [15].

C. Multi-object detection Model

As we mentioned in the introduction section, we employed a state-of-the-art multi-object detection model for multiple Cavendish bananas detection and ripeness classification, which is the YOLOv5 (You Only Look Once) Model [16] (Fig. 4). The model's complex architecture was built upon layers of neural network building blocks, which in general can be divided into three stages:

- Backbone: Contains convolutional neural networks responsible for assimilating and structuring image features across varied resolutions.
- Neck: A sequence of layers designed to combine essential image features that can be utilized in the prediction stage.
- Head: Processes the features derived from the neck by executing both bounding box and class prediction procedures.

Despite it's deep architecture, YOLOv5 employs clever strategies to reduce the model's parameters and size, ensuring both fast and precise prediction [17]. Since we plan to implement the multi-banana object classification model in mobile applications, a streamlined model size is crucial for efficient and fast inference on devices with limited resources. Moreover, the YOLOv5 model has been used in many case of fruit grading system and gives the best accuracy in many multi-object detection applications [18]. Therefore, we argue that YOLOv5 model was a suitable model to detect multiple bananas in an image and simultaneously predict their respective ripeness classes.

D. Training and Validation

To ensure the effectiveness and generalizability of the trained model, we randomly divided the augmented dataset to 85% for training and 15% for validation. The validation subset of the banana images was used to evaluate the model's performance on unseen data after training. The validation subset enables the assessment of the model's accuracy, precision, and recall metrics, providing insights into its ability to correctly classify multi-banana ripeness level across diverse conditions.

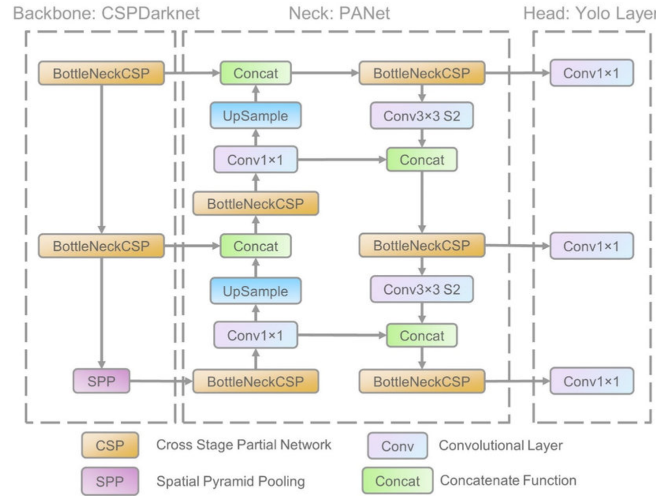


Fig. 4. Yolov5 Model Architecture [17]

E. Evaluation Metrics

The evaluation metrics used to assess the model's performance in our multi-banana ripeness classification case are the common evaluation metrics that are applied in general object detection model evaluation [19]. The metrics include mean average precision (mAP), precision, recall, and class precision. mAP is considered the most essential metric to indicate the performance of object detection and classification algorithms. mAP is acquired by first calculating the weighted mean of precision at each model's confidence threshold, and then averaging across all classification classes. Precision measures the proportion of correctly classified bananas, while recall assesses the model's ability to avoid false negatives. Finally, class precision evaluates the model's performance for specific ripeness categories.

Other than the previous metrics, three distinct types of loss functions are utilized in this model, which are box loss, object loss, and classification loss. The box loss quantifies the model's proficiency in pinpointing an object's central location and the adequacy of the predicted bounding box in encapsulating the object. Object loss serves as an indicator of the likelihood that a designated region of interest encompasses an object. Meanwhile, the classification loss provides insight into the model's capability to accurately ascertain the class of a specified object. These metrics and loss functions offer valuable insights for refining and improving the ripeness classification system.

F. Dataset Limitation

Initially, we plan to train the multi-object classification model to differentiate ripeness stage 1 to 7 of the local Cavendish bananas plus the overripe stage. However, since the banana samples are maturing during the distribution process from the plantation to the local shop, we cannot procure

TABLE I. Banana Classification Model metrics

Metrics	Value
mean Average Precision (mAP)	98.8
Precision	90.5
Recall	92.6
Average Precision for Stage 4	98.4
Average Precision for Stage 5	98.7
Average Precision for Stage 6	99.1
Average Precision for Stage 7	100
Average Precision for Overippen	98.7

specimens that are less than stage 4 of maturity. Therefore, our multi-object detection algorithm can only differentiate ripeness stage level 4 to 7 plus the overripe stage.

III. Results

To train the multi-object detection model for the ripeness classification of multiple Cavendish bananas, we utilized Roboflow platform [20] for rapid development. The proposed YOLOv5-based approach was given pre-trained COCO weights and then trained for 200 epochs using a training subset which is 85% of the randomly picked augmented banana images dataset. The training dataset consists of 2000 images of multiple Cavendish bananas at different ripeness stages from stage 4 to stage 7 and overripen stage. Training results show that the mAP begins to stabilize after 80^{th} iterations and reaches up to 98.8% as shown in Fig. 5. The class loss (Fig. 6a), object loss (Fig. 6b), and box loss (Fig. 6c) values also show decreasing trend over training epochs, indicating the convergence of the model after training.

Table I and Fig. 7 show the detailed evaluation metrics of the proposed multi-banana classification model during training and validation. Our proposed approach reaches 90.5% precision and 92.6% recall on the validation dataset. The average precision for each ripeness class also show promising results, reaching 98.4%, 98.7%, 99.1%, and 100% for ripeness stage 4, 5, 6, and 7, respectively. The performance of the trained YOLOv5 model on multi-banana detection and classification is even more significant than the four-layered CNN model that we used for single banana classification in the previous research [12] which only achieved 88.8% precision and 91.2% recall.

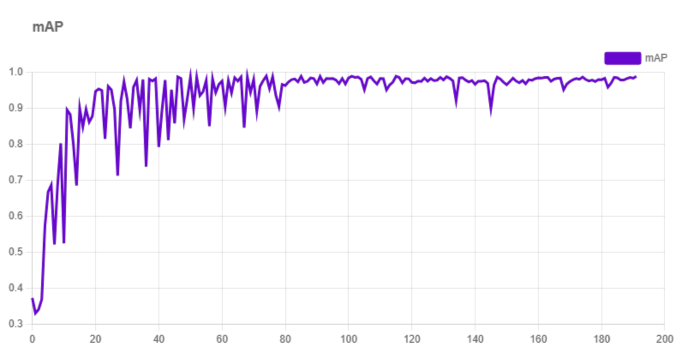


Fig. 5. Training Results

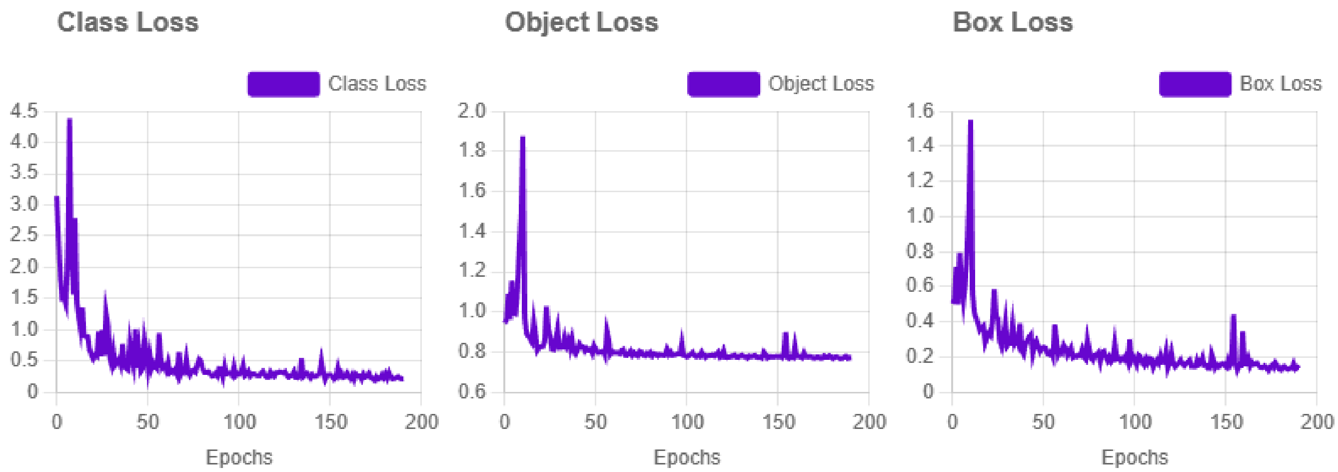
IV. Discussion, Limitation and Future Work

In this paper, we have accomplished the objective of developing a simultaneous multi-banana detection and classification algorithm tailored for a seven-class Cavendish banana ripeness levels based on YOLOv5 model. When evaluated with validation dataset, the model achieved 98.8% mean average precision, 90.5% precision, and 92.6% recall. To the best of our knowledge, our research is the first to attempt individual ripeness grading for multiple bananas in a single image. Therefore, there is no direct benchmark to the performance of our model. A general comparison to our previous research [12] which implemented four-layered CNN for single banana ripeness classification, shows that the YOLOv5 based model excels in performance by about 2% margin. The increase in metrics indicated that it is preferable to build on top of a proven and complex learning model such as YOLOv5 over custom shallow CNN that we implemented in the previous research. The 14 MB compact size and 10 ms fast inference characteristic of YOLOv5 model also make our multi-banana classification algorithm appropriate for mobile implementation along the banana distribution chain to converge the banana maturity grading system across all stakeholders.

However, it is vital to acknowledge the limitations of this research to encourage future improvements. As we mentioned in the Method section, the dataset collected for this research is limited to ripeness stage level 4 to 7 plus the overripen stage due to the natural ripening process of the banana samples during distribution. Additional data collection efforts from plantation is necessary to ensure the generalizability and robustness of the proposed approach. By addressing these limitations, future iterations of this system can provide even more accurate and reliable ripeness classification for Cavendish bananas for industry uses.

Another limitation of our study is the dependency on single-camera data collection. Using images from multiple cameras and environments could introduce additional challenges related to lighting and banana perspective variations [21]. Further improvement for the implementation of this system in the industry is the identification of partially visible bananas as can be found in banana bunches. To solve the problem, the bounding box of the object detection model should be narrowed to identify the exact geometry of the banana to make the classification system more precise. Other computer vision techniques, such as the Segment Anything Model (SAM) as recently used in the medical field [22] or CNN-Mask model as currently being experimented by [18], can complement our system to further enhance the capability of the fruit grading system.

Future research could also explore the feasibility of incorporating environmental information as additional input features to the multi-object detection model to track the ripening process over time. The method would enable more accurate predictions of the exact ripeness stage and facilitate better inventory management. Moreover, if the ripening rate of the banana can be accurately sensed and predicted, it would be



(a) Class losses (b) Object losses (c) Box losses

Fig. 6. Loss function evaluation metric results for each iteration

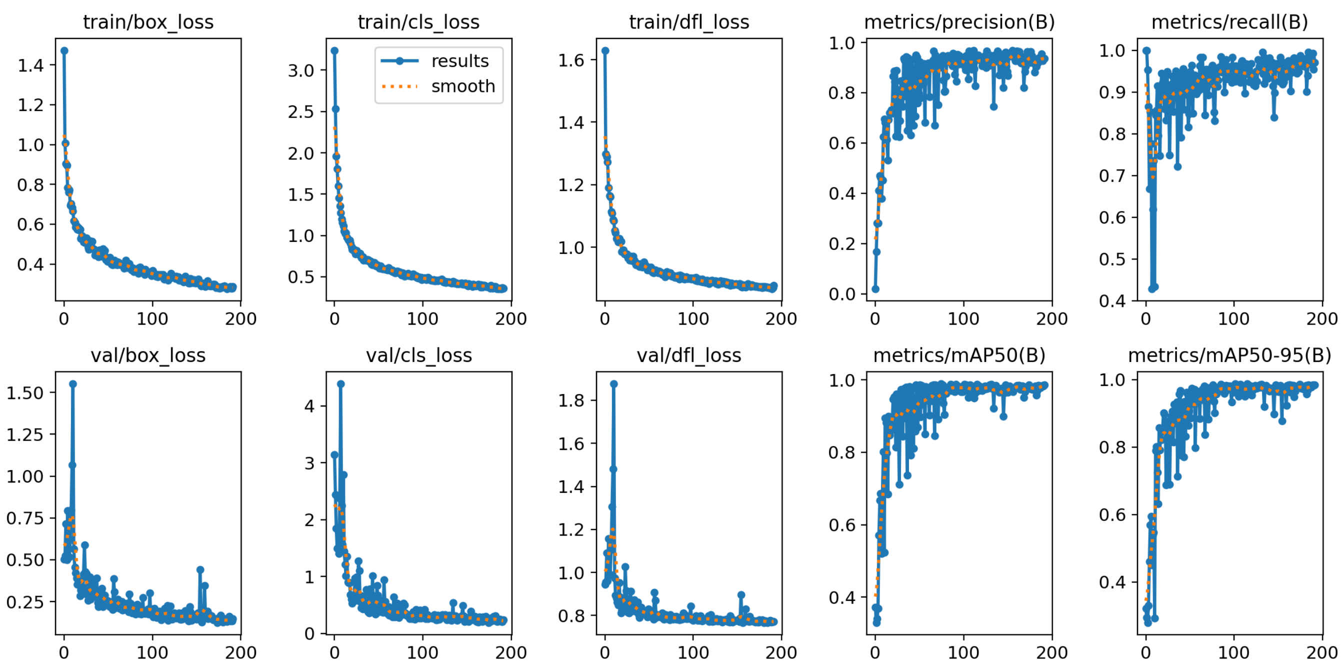


Fig. 7. Detailed Training and Validation Results

possible to apply a control technique to delay the ripening rate.

Overall, this research contributes to the advancement of agricultural technology by demonstrating the effectiveness of computer vision and deep learning techniques in fruit ripeness analysis and food quality assessment. The proposed approach opens avenues for future studies in this domain, encouraging further research and development to enhance the system's accuracy, scalability, and applicability.

V. Conclusion

Our study presents a novel approach for accurately classifying the ripeness of Cavendish bananas, a fruit of immense economic and nutritional significance. By leveraging the capabilities of the YOLOv5 (You Only Look Once) model, the study introduced a multi-object detection system that can simultaneously identify and classify the ripeness of multiple Cavendish bananas within images. The YOLOv5-based approach outperformed previous models, achieving a mean average precision of 98.8%, precision of 90.5%, and recall of 92.6% on the validation dataset. This research has the potential to revolutionize the banana industry by providing a rapid and reliable ripeness assessment process. However, further research is needed to enhance the model's generalizability and address limitations in the training data by conducting more comprehensive data collection. Furthermore, incorporating environmental data as additional input features could refine ripeness predictions. Exploring other computer vision techniques, such as the Segment Anything Model (SAM) or CNN-Mask model, could also enhance the system's capability. Overall, this study contributes to the advancement of fruit ripeness analysis and demonstrates the potential of computer vision in agricultural applications. By providing a reliable and automated solution for the evaluation of ripeness, it paves the way for future studies and innovations in the domain of food quality evaluation.

ACKNOWLEDGMENT

We extend our sincere appreciation to the head of our research center, Yusuf Nur Wijayanto, for his exceptional leadership, unwavering support, and guidance throughout this research project. We are also grateful to our esteemed research partner, Prof. Dr. Fenny Dwivany from the School of Life Science and Technology, Institut Teknologi Bandung, for her invaluable contributions and cooperation, which have enriched the scope and quality of this study.

REFERENCES

[1] FAO, "International Trade Banana: Market Review Preliminary Results 2021," United Nations, Rome, Tech. Rep., 2022.

[2] C. K. Saha, M. K. Ahamed, M. S. Hosen, R. Nandi, and M. Kabir, "Post-harvest Losses of Banana in Fresh Produce Marketing Chain in Tangail District of Bangladesh." *Journal of the Bangladesh Agricultural University*, vol. 19, no. 3, pp. 389–397, 2021. [Online]. Available: https://jbau.bau.edu.bd/index.php/home/article/view/32

[3] P. Baglat, A. Hayat, F. Mendonça, A. Gupta, S. S. Mostafa, and F. Morgado-Dias, "Non-Destructive Banana Ripeness Detection Using Shallow and Deep Learning: A Systematic Review," *Sensors*, vol. 23, no. 2, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/738

[4] M. Knott, F. Perez-Cruz, and T. Defraeye, "Facilitated machine learning for image-based fruit quality assessment," *Journal of Food Engineering*, vol. 345, no. 2022, p. 111401, 2023.

[5] J. Prihantoro, D. R. Anbiya, G. Prihantoro, E. Nashrullah, Z. Cahya, and Riza, "Comparison of Machine Learning Techniques for Grade Classification of Exported Cavendish Bananas," in *2023 IEEE International Conference on Data and Software Engineering (ICoDSE)*. IEEE, sep 2023, pp. 37–42. [Online]. Available: https://ieeexplore.ieee.org/document/10291592/

[6] A. D. M. Africa, "Ripe Fruit Detection and Classification using Machine Learning," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1845–1849, 2020.

[7] N. Saranya, K. Srinivasan, and S. K. Kumar, "Banana ripeness stage identification: a deep learning approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 8, pp. 4033–4039, 2021. [Online]. Available: https://doi.org/10.1007/s12652-021-03267-w

[8] J. Ni, J. Gao, L. Deng, and Z. Han, "Monitoring the Change Process of Banana Freshness by GoogLeNet," *IEEE Access*, vol. 8, pp. 228 369–228 376, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9296756/

[9] F. M. A. Mazen and A. A. Nashat, "Ripeness Classification of Bananas Using an Artificial Neural Network," *Arabian Journal for Science and Engineering*, vol. 44, no. 8, pp. 6901–6910, aug 2019. [Online]. Available: https://doi.org/10.1007/s13369-018-03695-5 http://link.springer.com/10.1007/s13369-018-03695-5

[10] R. E. Saragih and A. W. Emanuel, "Banana Ripeness Classification Based on Deep Learning using Convolutional Neural Network," in *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, no. April. IEEE, apr 2021, pp. 85–89. [Online]. Available: https://ieeexplore.ieee.org/document/9431928/

[11] Y. Zhang, J. Lian, M. Fan, and Y. Zheng, "Deep indicator for fine-grained classification of banana's ripening stages," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 46, dec 2018. [Online]. Available: https://jivp-eurasipjournals.springeropen.com/articles/10.1186/s13640-018-0284-8

[12] Z. Cahya, D. Cahya, T. Nugroho, A. Zuhri, and W. Agusta, "CNN Model with Parameter Optimisation for Fine-Grained Banana Ripening Stage Classification," in *Proceedings of the 2022 International Conference on Computer, Control, Informatics and Its Applications*, ser. IC3INA '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 90–94. [Online]. Available: https://doi.org/10.1145/3575882.3575900

[13] R. Phoophuangpairoj, T. Ngoenrungrueang, and S. Audomsin, "Ripeness Classification of a Bunch of Bananas Using a CNN," in *2023 International Electrical Engineering Congress (iEECON)*. IEEE, mar 2023, pp. 180–183. [Online]. Available: https://ieeexplore.ieee.org/document/10126585/

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0

[15] F. Harby, A. Thaljaoui, D. Nayab, S. Aladhadh, S. EL Khediri, and R. Ullah Khan, "Data Augmentation and Random Multi-Model Deep Learning for Data Classification," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 5191–5207, 2023. [Online]. Available: https://www.techscience.com/cmc/v74n3/50986

[16] H. Liu, F. Sun, J. Gu, and L. Deng, "Sf-yolov5: A lightweight small object detection algorithm based on improved feature fusion mode," *Sensors*, vol. 22, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/15/5817

[17] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, p. 217, 2021.

[18] K. Goyal, P. Kumar, and K. Verma, "AI-based fruit identification and quality detection system," *Multimedia Tools and Applications*, vol. 82, no. 16, pp. 24 573–24 604, jul 2023. [Online]. Available: https://link.springer.com/10.1007/s11042-022-14188-x

[19] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/3/279

[20] B. Dwyer, J. Nelson, J. Solawetz, et al., "Roboflow," 2022. [Online]. Available: https://roboflow.com

[21] Q. Hao, X. Guo, and F. Yang, "Fast Recognition Method for Multiple Apple Targets in Complex Occlusion Environment Based on Improved YOLOv5," *Journal of Sensors*, vol. 2023, p. 3609541, 2023. [Online]. Available: https://doi.org/10.1155/2023/3609541

[22] P. Shi, J. Qiu, S. M. D. Abaxi, H. Wei, F. P.-W. Lo, and W. Yuan, "Generalist vision foundation models for medical imaging: A case study of segment anything model on zero-shot medical segmentation," *Diagnostics*, vol. 13, no. 11, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/11/1947