Advancing Fruit Recognition: Multiclass Detection Enabled through Machine Vision Technology

Pradeep Kumar[1*],, Arjun Chouriya[2],

[1*]PhD Scholar, Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute, New Delhi 110012.
[2]Ph.D Scholar, Indian Institute of Technology, Kharagpur, 721302

*Corresponding author email address: rajpootpradeep112233@gmail.com

Abstract

One of the key elements in predicting yield is the real-time detection of fruits and vegetables. Conventional detection techniques encounter constraints in precisely discerning distinct growth stages, primarily attributable to pronounced occlusion stemming from proximal foliage, substantial intermingling among adjacent fruits, dissimilarities in dimensions, pigmentation, aggregation density, and diverse developmental attributes. An improved YOLO-v3 model is suggested for recognising apples in orchards with varying light, complex backdrops, overlapping apples, branches, and leaves in order to detect the fruits at various development phases. Images of young, growing, and ripe apples are the first things that are gathered. Subsequently, rotation-based transformations, color equilibrium adjustments, luminance manipulations, and image blurring procedures are employed to augment the visual quality of these photographs. In recent fruit detection models, both Faster R-CNN with VGG16 net, and original YOLO-v3-dense are outperformed by an improved model YOLO-v3-dense. In a scenario including a complicated orchard, the Dense-YOLOv4 model has been used to identify several mango growth phases with a high degree of occultation. By concluding all above points, it is suggested that Dense-YOLOv4 model is one of most accurate model to detect the different types of fruits and vegetables.

Introduction

India is renowned for its diverse climatic conditions that facilitate the cultivation of a wide variety of fresh fruits and vegetables. In global fruit and vegetable production, India holds the second position, trailing only behind China. Notably, India contributes approximately 15% to the global fruit output **[6,1].** The cultivation landscape spans 6.66 million hectares, yielding a substantial 102.08 million metric tonnes of fruits **[3, 14].** Nevertheless, the agricultural sector grapples with the significant impact of plant diseases and pests, leading to ecological and yield-related losses. Addressing these concerns, the early detection and prevention of diverse plant diseases have emerged as pivotal strategies within agricultural technology, particularly for viable farms and orchards **[15].** Traditional methodologies, centered around manual visual inspections for disease identification, exhibit inefficiency and protracted timelines, subsequently inflating operational expenses. In recent times, the realm of precision agriculture has been revolutionized by cutting-edge advancements in computer vision **[19,5].** This revolution has seamlessly integrated disease detection protocols into crop health monitoring practices, markedly enhancing the efficiency of disease identification and augmenting overall crop yield. The timely recognition and mitigation of plant diseases hold paramount significance in safeguarding crop health and optimizing harvests by curtailing growth irregularities **[8]**. Such interventions also curtail the need for extensive pesticide application, aligning with the aspiration for environmentally friendly crop production. In light of these imperatives, the deployment of automated plant disease detection, leveraging diverse machine learning algorithms, has emerged as an efficacious approach within the domain of precision agriculture.

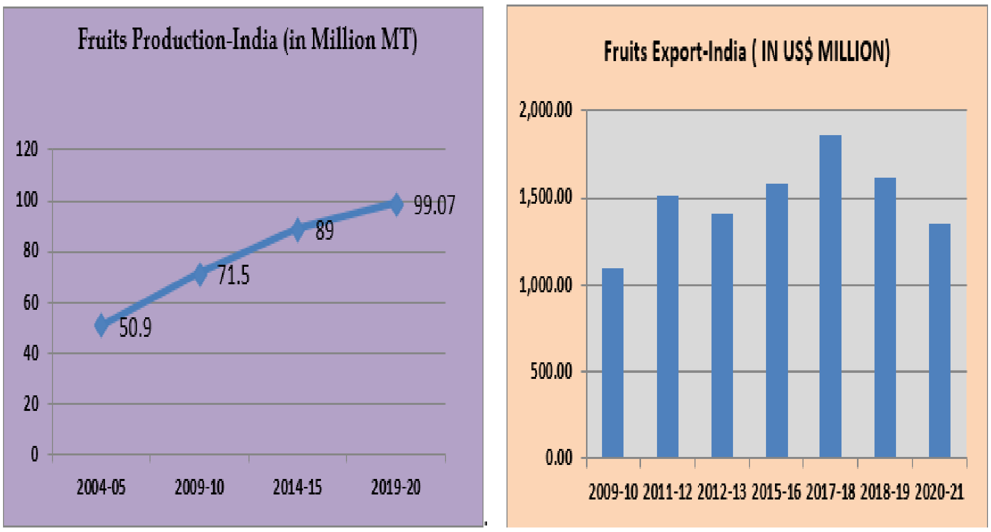


Figure 1: Fruits production in India and fruits export from India (Agriculture Export Policy 2020-21[2]).

Challenges in Manual Detection

Agricultural labor is inherently seasonal, confining workforce engagement to select months. This intermittency compels laborers to seek permanent roles in non-agricultural sectors, attributing to the burgeoning wage rates outside of agriculture **[9]**. Manual inspection methods are associated with prolonged time investments and escalated production expenses due to the substantial labor costs incurred. The identification of distinct fruit growth stages necessitates a skilled and experienced labor force. Inexperienced workers invariably introduce avoidable errors, thereby undermining productivity

Figure 1: Challenges in manual detection of fruits

Advantages of Machine Vision System Over Manual Detection

Progressive strides in precision agriculture and information technology have catalyzed the fusion of robotics, crop imaging, computer vision, and object detection. These synergistic components facilitate accurate data acquisition, pivotal for assessing crop progress and monitoring overall health. The discernment of distinct agricultural growth stages is pivotal for prognosticating future yields, facilitating smart sprayer systems, and orchestrating self-governing pesticide-dispensing robots across expansive farms and orchards **[10].** However, the challenges of achieving precise target object detection persist due to factors like color and texture resemblances, intricate backgrounds, overlapping entities due to dense distribution, variable illumination across sprawling terrains, and sundry other variables. These complexities underscore the need for enhanced accuracy in target object detection, a domain where machine vision systems excel. Furthermore, this paradigm expedites operations compared to manual counterparts, concurrently ameliorating growth irregularities via prompt detection. Machine vision system can work on a variety of algorithms depending on the specific task they are designed for. Machine vision involves using cameras or other imaging sensors to acquire and process visual information in order to make automated decisions or perform specific tasks **[4].**

Algorithm

The term "algorithm" denotes a systematic assemblage of rules and instructions employed for computational or problem-solving endeavors. It encapsulates a sequential delineation of steps that dictate the execution of tasks to achieve predetermined outcomes.

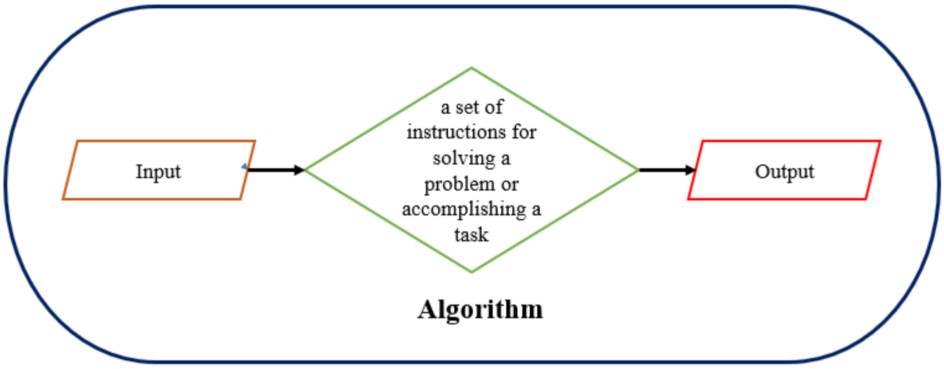


Figure 2: Algorithm definition

Characteristics of an Algorithm:

• **Precision and Clarity:** An algorithm necessitates precision and clarity. Each step within it must be distinctly defined, leaving no room for ambiguity, and should lead to a singular interpretation.

• **Precise Input Specification:** In the event that an algorithm involves taking inputs, these inputs must be precisely and unambiguously specified.

• **Explicit Output Specification:** The algorithm must unequivocally outline the nature of its output, providing explicit details of what will be generated.

• **Finiteness:** Crucially, the algorithm must be characterized by finiteness. It should steer clear of infinite loops or comparable scenarios that could lead to perpetual execution.

• **Practicality:** An algorithm must exhibit practicality, simplicity, and universality. It should be feasible to execute with available resources, devoid of reliance on speculative technologies or future developments.

• **Language Neutrality:** The algorithm's design must transcend language barriers. It ought to consist of plain, universally understandable instructions that can be implemented across languages, while still yielding consistent and anticipated outcomes.

Different Algorithm used in fruits detection

There are several ways to understand the description of recognition (detection and classification):

(i) Identification of a fruit (distinguishing between a fruit and an item, such as a leaf from a backdrop).

(ii) Classification of the fruit classes (e.g., orange and tangelo).

(iii) Differentiating between a variety of fruit species

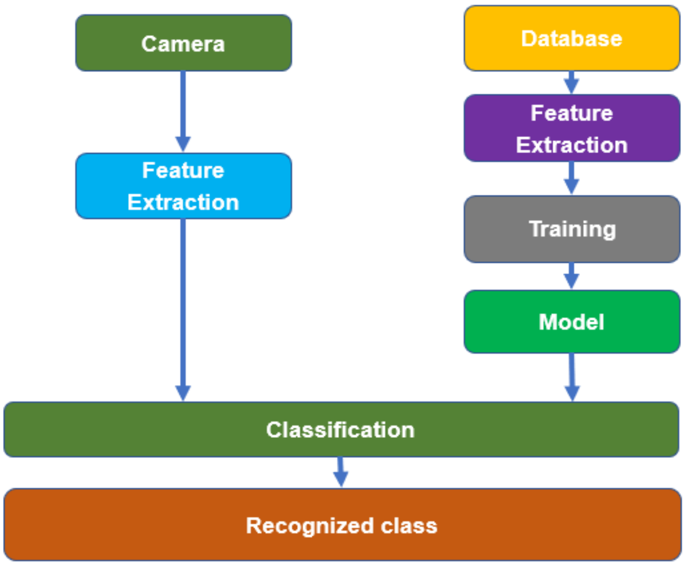


Figure 4: Basic architecture for fruit detection

To address the challenge of identifying suitable fruit species and types, it's essential to first understand the complexity of the task. The classification of fruits is complicated due to their vast variety, leading to prominent differences in shapes, colors, and textures. Moreover, compounded by limited image scope due to factors such as lighting, angles, and distances during image capture, the result is often unclear images **[17].** Additionally, objects can be partially or entirely hidden, adding to the issue. These difficulties have hindered the practical adoption of multi-class automated fruit classification systems in real-world scenarios.

Object detection algorithms are categorically divided into two groups depending on the number of instances an identical input image undergoes within a network.

Single-shot object detection

Single-shot object detection involves making predictions about object presence and locations in an image through a singular traversal of the input image. This approach exhibits computational efficiency by handling the entire image in a singular pass. However, compared to alternative techniques, the accuracy of single-shot object detection is generally lower, particularly concerning the detection of diminutive objects **[21]**. This methodology proves advantageous for real-time object detection in resource-constrained scenarios. YOLO, a representative of single-shot detectors, employs a fully convolutional neural network (CNN) to process images. The subsequent section will provide an in-depth exploration of the YOLO model.

Two-shot object detection

The dual-shot object detection technique involves employing two sequential scans of the input image to infer object presence and spatial positioning. The first pass is used to generating a series of proposals outlining potential object locations, followed by a subsequent pass aimed at refining these proposals and making final predictions. While offering heightened accuracy in comparison to single-shot object detection, this approach does entail greater computational demands. The selection between single-shot and two-shot object detection hinges upon the specific prerequisites and limitations of the application at hand **[12].** Generally, real-time scenarios are better served by single-shot object detection, whereas instances prioritizing precision are better suited for the two-shot approach.

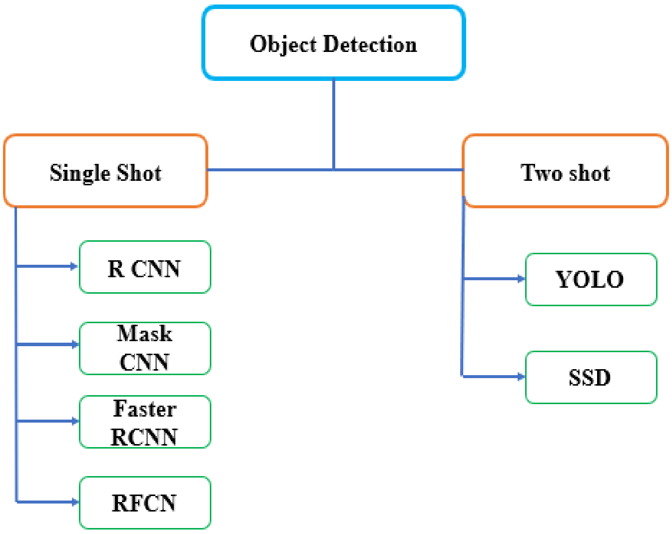


Figure 3: One stage and two stage detector Algorithm

The algorithms which are being used frequently for fruits detection are CNN and YOLO.

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) constitute a variant of neural network architecture facilitating the acquisition of enhanced image feature representations. However, real-time object detection capabilities are not inherently intrinsic to CNNs. CNNs have emerged as a pivotal framework for scholarly exploration in domains encompassing object classification and image comprehension. The resilience of CNNs stems from their innate capacity to autonomously extract salient attributes from input images **[7]**.

You Only Look Once (YOLO)

YOLO is a neural network-based algorithm designed for instant object detection. It works by utilizing regression to predict both object classes and bounding boxes for the entire image in a single algorithm run **[16]**. Notably, this approach necessitates just one pass through a neural network to accomplish object detection.

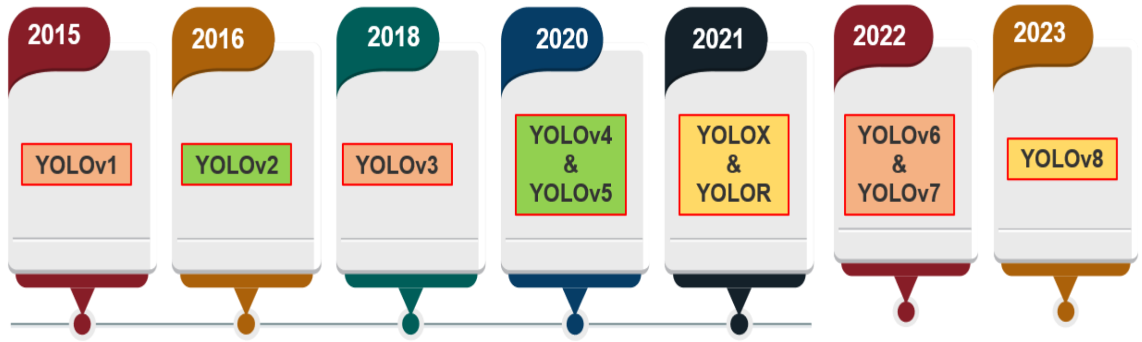
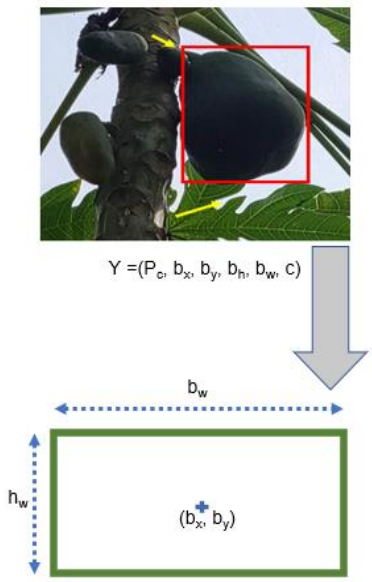


Figure 6: Timeline of evolution of YOLO algorithm

Working Principle of Fruit detection Algorithm

The following principles are used by the YOLO algorithm.



- Bounding box regression
- Intersection Over Union (IOU)

Bounding box regression

A bounding box is a defined border that emphasizes an object within an image.

- Each bounding box within the image includes the subsequent attributes:
- Width (b_w)
- Height (b_h)
- Class (This is represented by the letter c).
- Bounding box center (b_x, b_y)

Intersection over Union (IOU)

Intersection over Union (IOU) is a concept in object detection that explains the extent of overlap between boxes. YOLO employs IOU to generate an output box that accurately encloses the objects.

This mechanism eliminates bounding boxes that are not equal to the real box.

$$\text{IOU} = \frac{Soverlap}{Sunion}$$

Where, $S_{overlap}$ is the area of intersection of the predicted bounding box and the actual bounding box,

S_{union} is the area of the union of the both bounding boxes.

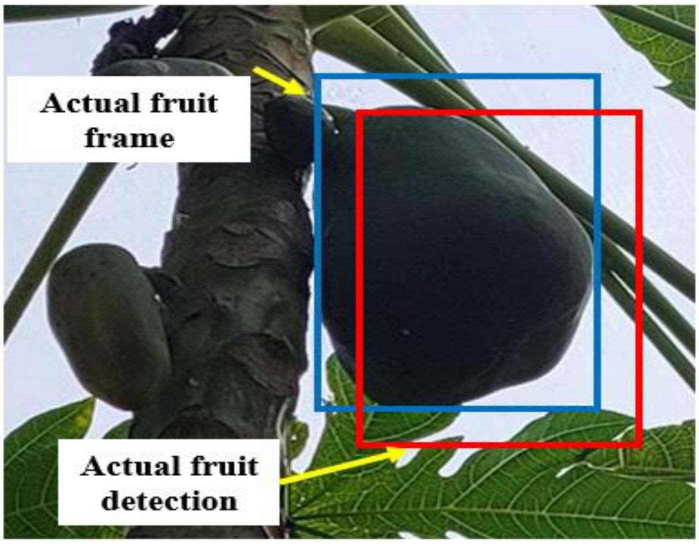


Figure 5: Actual fruit frame and Detected fruit frame

Different indices used in fruit detection algorithm

In the realm of deep learning-driven object detection models, pivotal statistical measures are employed for assessment, encompassing matrices like intersection over union (IoU), precision (P), recall (R), F-1 score, average precision (AP), and mean average precision (mAP). Within YOLOv4, an evaluation metric termed IoU emerges as a standard tool to gauge the accuracy of object detection. IoU is invariant to scale and is utilized to quantify the model's performance efficiency. IoU operates by quantifying the ratio of the overlapping area between the bounding box prediction generated by the model and the actual bounding area of the object **[11].** This comparison serves as a reflection of the model's efficacy and performance. The mathematical formulation of IoU is expressed as

$$\text{IOU} = \frac{Aoverlap}{Aunion}$$

where $A_{overlap}$ is described as the region where the true bounding box of the object and the model's predicted bounding box intersect. A_{union}, on the other hand, is the union of the aforementioned bounding boxes. If IoU is larger than 0.5 for binary classification, the classified object class can be described as true positive (TP). IoU values below 0.5 can be classified as false positives (FP) for the appropriate class. Using the definitions of TP, FP, and FN, the performance variables P and R can be stated as follows

$$\text{Precision (P)} = \frac{TP}{TP+FP} \times 100\ \%$$

Where, TP = True positive

FP = False positive

FN = False Negative

According to the above equation, higher P denotes models' enhanced ability to discriminate between negative datasets, and higher R denotes models' enhanced ability to detect positive datasets. Using the above Equation to determine test accuracy, the F1 score can be defined as follows:

$$F_1\text{-score} = \frac{2 \times precision \times recall}{precision + recall} \times 100\ \%$$

In order to balance the precision and recall of the model, the F1 score is evaluated as an indicator for integrating the mean of the precision and recall. In general, a model with a higher F1 score is more robust than one with a lower one. In a broader context, the average precision (AP) corresponds to the area under a precision-recall curve (PR-curve), and this can be formulated as follows

$$P_{\text{average}} = \sum_{j=1}^{N\ (class)} precision(j) \times recall(j) \times 100\%$$

When AP is higher, there is a larger area under the PR curve, which indicates that the object class can be predicted more accurately, whereas mAP is the average of all APs, which can be expressed as

$$\text{Mean Average Precision (mAP)} = \frac{Paverage}{N\ (class)}$$

The confidence scores can be expressed as

Confidence = pr(object) × IoUtruth pred ∨ pr(object) ∈ 0, 1

If the target class falls within the YOLO grid, pr(object) = 1 is prescribed; otherwise, pr(object) = 0. The IoU truth prediction describes the coincidence between the reference and the predicted bounding box. Here, IoU is the intersection over union. When the specified class is identified within the grid, the value of "pr(object)" signifies the precision of the bounding box prediction **[13].** For the final bounding box, the best predictions from each of these scales are filtered using the non-maximum suppression (NMS) algorithm.

Comparison of different algorithm used in fruit detection

The study involves a comparative analysis of detection outcomes concerning the developmental stages of mangoes using YOLOv3, YOLOv4, and Dense-YOLOv4 algorithms. To enhance the precision of bounding box representations, the investigation focuses on four distinct growth phases: budding, early growth, intermediate growth, and full maturation. The bounding box classes align with these growth stages. Examination of the detection results reveals that the novel Dense-YOLOv4 algorithm consistently exhibits heightened accuracy in bounding box predictions across all identified growth phases, surpassing the performance of both YOLOv3 and YOLOv4 **[20].** The initial budding phase presents a particular challenge due to the discrete yet densely clustered appearance of mangoes. The intricate textural resemblance among neighbouring buds complicates the individual object detection task. However, the outcomes of Dense-YOLOv4 exhibit a pronounced enhancement in detection precision and a reduction in the count of undetected objects, notably outperforming the conventional YOLOv3 and YOLOv4 approaches in this context. The comparison of different algorithm is given in Table-1.

Table.1-Comparison of P, R, F1-Score, *mAP*, and detection speed (FPS) between Dense-YOLOv4 and other state-of-art models

Model	P(%)	R(%)	F1-Score (%)	*mAP (%)*	Det. Time (ms)	FPS
Faster R-CNN	53.64	66.27	65.36	59.17	44.41	22.59
YOLO-v4	83.51	82.77	83.14	91.47	20.81	49.82
Dense-YOLO-v4	91.45	95.87	93.61	96.20	22.62	44.20
YOLO-v3	75.78	85.57	80.38	89.19	23.25	43.13
Mask R-CNN	69.27	74.58	71.82	73.40	33.71	29.67

(Roy and Bhaduri 2022 [18])

Conclusions

In essence, the YOLO (You Only Look Once) algorithm, which employs neural networks for instantaneous object detection, has exhibited notable progress in its domain. Specifically, the YOLO-v3-Dense iteration has showcased superior performance relative to its precursor, YOLO-v3. This advancement becomes especially evident when comparing it against the leading Faster R-CNN model featuring the VGG16 architecture; the Dense-YOLO-v3 model has established its supremacy in discerning fruits. Moreover, recent empirical analyses have brought to light the considerably heightened efficacy of the Dense-YOLO-v4 algorithm when compared with the original YOLO-v4 model, with a pronounced emphasis on fruit detection precision and overall accuracy. These collective findings highlight the continuous evolutionary trajectory of the YOLO framework, culminating in progressively refined versions that establish fresh benchmarks in real-time object detection, particularly in scenarios involving the meticulous identification of fruits. This conveys the algorithm's steady march towards achieving superior object detection outcomes in intricate, real-world contexts.

References

[1] Anonymous, (2020). Fruits and Vegetables, Agricultural and Processed Food Products Export Development Authority (APEDA), Ministry of Commerce & Industry, Government of India.

[2] Anonymous, (2020). India registered excellent growth in Agriculture Exports during 2020-21, Ministry of Commerce & Industry, Government of India.

[3] Anonymous, (2021). Final Estimates of 2020-21 and First Advance Estimates of 2021-22 of Area and Production of Horticultural Crops, Ministry of Agriculture & Farmer Welfare, Government of India.

[4] Botta, A., Cavallone, P., Baglieri, L., Colucci, G., Tagliavini, L., & Quaglia, G. (2022). A review of robots, perception, and tasks in precision agriculture. *Applied Mechanics*, *3*(3), 830-854.

[5] Dixit, S., Bohre, K., Singh, Y., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2023). A Comprehensive review on AI-enabled models for Parkinson’s disease diagnosis. *Electronics*, *12*(4), 783.

[6] Gnanavel, S., Manohar, S., Sridhar, K. E., Sokkanarayanan, S., & Sathiyanarayanan, M. (2019, December). Quality detection of fresh fruits and vegetables to improve horticulture and agro-industries. In *2019 International Conference on contemporary Computing and Informatics (IC3I)* (pp. 268-272). IEEE.

[7] Gray, P. C., Fleishman, A. B., Klein, D. J., McKown, M. W., Bezy, V. S., Lohmann, K. J., & Johnston, D. W. (2019). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *10*(3), 345-355.

[8] Haldhar, S. M., Saha, R. K., & Nagesh, M. (2021). Souvenir-cum-Abstract Book: National Conference on Priorities in Crop Protection for Sustainable Agriculture.

[9] https://www.fao.org/3/cb9479en/online/sofa-2022/future-agrifood-workforce.html

[10] Hu, N., Su, D., Wang, S., Nyamsuren, P., Qiao, Y., Jiang, Y., & Cai, Y. (2022). LettuceTrack: Detection and tracking of lettuce for robotic precision spray in agriculture. *Frontiers in Plant Science*, *13*, 1003243.

[11] Koech, K. E. (2020). Object detection metrics with worked example: AP, mAP, AP50 among other metrics explained with an example. *Towards Data Science*.

[12] Kundu, R. (2023). YOLO Algorithm for Object Detection Explained [+ Examples].

[13] Mercaldo, F., Brunese, L., Martinelli, F., Santone, A., & Cesarelli, M. (2023). Object Detection for Brain Cancer Detection and Localization. *Applied Sciences*, *13*(16), 9158.

[14] Mistry, I., & Gundawar, P. (2022). Consumer Awareness and Perception towards the New Concepts and Technology in Agriculture Industry.

[15] Orchi, H., Sadik, M., & Khaldoun, M. (2021). On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey. *Agriculture*, *12*(1), 9.

[16] Rani, E. (2021). LittleYOLO-SPP: A delicate real-time vehicle detection algorithm. *Optik*, *225*, 165818.

[17] Rizzo, M., Marcuzzo, M., Zangari, A., Gasparetto, A., & Albarelli, A. (2023). Fruit ripeness classification: A survey. *Artificial Intelligence in Agriculture*.

[18] Roy, A. M., and Bhaduri, J. (2022). Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. *Computers and Electronics in Agriculture*, *193*, 106694.

[19] Shin, J., Mahmud, M. S., Rehman, T. U., Ravichandran, P., Heung, B., & Chang, Y. K. (2022). Trends and prospect of machine vision technology for stresses and diseases detection in precision agriculture. *AgriEngineering*, *5*(1), 20-39.

[20] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., and Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, *157*, 417-426.

[21] Tsang, S. H. (2018). Review: SSD—single shot detector (object detection). *medium. com. https://towardsdatascience. com/review-ssd-single-shotdetector-object-detection-851a94607d11 (accessed August, 25, 2023)*.