



Artificial Intelligence based Precise Disease Detection in Soybean using Real Time Object Detectors

Aditya Kamalakara Kanade¹, M.P. Potdar², Gurupada Balol³, Nagesh Rathod¹, Pooja⁴, S.N. Huligol⁴, R. Channakeshava⁴, K.N. Vijaykumar⁵

10.18805/LR-5431

ABSTRACT

Background: Soybean, a crucial oilseed crop in India, is susceptible to various diseases, including Yellow Mosaic Disease caused by Mungbean Yellow Mosaic Virus (MYMV). Early detection of plant leaf diseases is a major necessity for controlling the spread of infections and enhancing the quality of food crop. Therefore, this study aims to evaluate the efficacy of YOLOv8 for detecting Mungbean Yellow Mosaic Virus in soybean crops.

Methods: Images were captured using 48 MP mobile camera from the experimental fields of Main Agricultural Research Station, University of Agricultural Sciences, Dharwad during 2024 summer season. The captured images were annotated, augmented using various transformations, resized and normalized. As a result, a dataset of 4480 annotated images of MYMV-infected soybean plants was created. The dataset was randomly split into training and validation set. Different variants of YOLOv8 models were trained and evaluated based on precision, recall, mean Average Precision (mAP) and inference time.

Result: Among the variants, YOLOv8x demonstrated superior performance, with a precision of 98.1%. Whereas the recall (78.6%), mAP@0.5 (85.3%) and F1-score (86.9%) were higher in YOLOv8s. YOLOv8n boasted the lowest inference time of 3.6 milliseconds and the least number of model parameters (3.1 millions). YOLOv8s shows promise as an efficient and accurate tool for MYMV detection in soybean crop. Its low inference time and high precision make it suitable for real-world deployment, aiding in early disease detection and precise management of the disease. This study contributes to advancing disease detection techniques in agriculture, facilitating early intervention and minimizing crop losses for sustainable agricultural practices.

Key words: Disease detection, MYMV, Soybean, YOLOv8.

INTRODUCTION

Soybean stands as one of the most critical oilseed crops in India, with an annual production reaching 13.98 mt. Characterized by its luxuriant growth and lush green, soft, succulent and nutrient-dense foliage, it is susceptible to numerous fungal, bacterial and viral diseases throughout its growth cycle, from sowing to harvesting stages. More than 50 viruses have been identified as causing various diseases in soybean crops (Mandhare and Gawade, 2010). Among these, Mungbean Yellow Mosaic India Virus (MYMIV) and Mungbean Yellow Mosaic Virus (MYMV) are the most prevalent and frequently infects soybean crop causing a disease collectively called as Yellow Mosaic Disease (YMD). In India, YMD cause yield losses varying from 10 to 88% in soybean annually (Singh *et al.*, 2013; Rohit *et al.*, 2023).

The prompt identification of diseases holds paramount importance in agriculture. It enables farmers to swiftly implement targeted treatments or isolate affected plants, thereby curbing the spread of disease and safeguarding their crops. Conversely, misdiagnosis can lead to inappropriate treatments, potentially harming healthy plants. Hence, accurate and rapid disease detection is crucial for effective prevention and management (Mathew and Mahesh, 2022). Traditional manual disease detection methods rely heavily on farmers' observational skills or consultation with experts, which are slow, inefficient, costly, subjective and prone to inaccuracies (Heltin *et al.*, 2019).

¹Department of Agronomy, University of Agricultural Sciences, Dharwad-580 005, Karnataka, India.

²AICRP (Dry Land Agriculture), University of Agricultural Sciences, Dharwad-580 005, Karnataka, India.

³AICRP (Groundnut), University of Agricultural Sciences, Dharwad-580 005, Karnataka, India.

⁴AICRP on Soybean, University of Agricultural Sciences, Dharwad-580 005, Karnataka, India.

⁵ICAR-Indian Agricultural Research Institute, Regional Research Centre, UAS Campus, Dharwad-580 005, Karnataka, India.

Corresponding Author: M.P. Potdar, AICRP (Dry Land Agriculture), University of Agricultural Sciences, Dharwad-580 005, Karnataka, India. Email: potdarp@uasd.in

How to cite this article: Kanade, A.K., Potdar, M.P. Balol, G., Rathod, N., Pooja, Huligol, S.N., Channakeshava, R. and Vijaykumar, K.N. (2025). Artificial Intelligence Based precise Disease Detection in Soybean Using Real Time Object Detectors. *Legume Research*. 48(11): 1878-1883. doi: 10.18805/LR-5431.

Submitted: 10-10-2024 **Accepted:** 04-01-2025 **Online:** 08-02-2025

Integrating deep learning technology into agricultural disease identification streamlines this process, reducing workload and significantly enhancing disease recognition accuracy. This integration is pivotal for precise crop disease identification (Yang *et al.*, 2024). With the advancement of computing technology, machine learning and image

processing have emerged as powerful tools for automating the detection and identification of plant diseases. Some studies evaluated performance of sparse recovery (Bazzi *et al.*, 2016) and multi-snapshot compressed sensing algorithms (Bazzi *et al.*, 2017). Machine vision plays a crucial role in the automatic diagnosis of plant diseases, offering efficient solutions for agricultural disease management (Zhao *et al.*, 2016).

YOLO is a series of real time object detector, which leverage CNN for single shot object detection and has been used for disease detection in different crops (Shill and Rahman, 2021; Nasution and Kartika, 2022). Uoc *et al.* (2022) focused on disease detection in cucumber plants using the YOLOv4 network for leaf image analysis. The authors achieved impressive accuracy rates exceeding 80% with a dataset comprising over 7000 images. Soeb *et al.* (2023) validated the detection and identification results of the YOLOv7 approach using prominent statistical metrics such as detection accuracy, precision, recall, mAP value and F1-score, which yielded values of 97.3%, 96.7%, 96.4%, 98.2% and 0.965, respectively. The experimental findings indicate that YOLOv7 outperforms existing target detection and identification networks, including CNN, Deep CNN, DNN, AX-Retina Net, improved DCNN, YOLOv5 and Multi-objective image segmentation, particularly for detecting tea leaf diseases in natural scene images. Sangaiah *et al.* (2024) introduced the enhanced UAV Tiny YOLO Rice (UAV T-yolo-Rice) network, which achieved a testing mean average precision (mAP) of 86% after training on the proposed dataset of rice leaf diseases.

Building upon this existing body of research, this study investigates the efficacy of YOLOv8 for disease detection in soybean plants. The significance of YOLOv8 lies in its state-of-the-art object detection capabilities, known for their accuracy, efficiency and scalability. This study is guided by the hypothesis that YOLOv8, particularly its smaller variants such as YOLOv8n, can efficiently detect Yellow Mosaic Virus (MYMV) in soybean with high accuracy and low computational complexity. This hypothesis is rooted in the expectation that the lightweight architecture and low inference time of YOLOv8n make it particularly suitable for real-time agricultural applications. Moreover, there is a lack of disease detection datasets specifically related to Indian production systems. By testing YOLOv8 on MYMV (*Mungbean Yellow Mosaic Virus*) disease, this research aims to contribute to the advancement of disease detection in agriculture, ultimately facilitating early intervention and minimizing crop losses for improved agricultural productivity and sustainability. The specific objectives of the study are:

- a. To build disease detection dataset in soybean for detection of MYMV infected plants.
- b. To train a robust real time YOLOv8 models and compare the performance of different variants of YOLOv8.

MATERIALS AND METHODS

Image acquisitions and dataset preparation

Images for the detection of MYMV were acquired using a mobile camera (Samsung Galaxy F22) equipped with a 50-megapixel sensor. The data collection was conducted at 35 Days after crop sowing in the month of February, 2024 during daytime under natural lighting conditions to ensure optimal visibility and image clarity. Each image was captured at a height ranging from 40-50 cm above the soybean crop height to encompass a comprehensive view of the field. All images were saved in the JPEG format. An example of the images captured during the experiment is depicted in Fig 1.

Bounding box annotations were drawn around the instances of MYMV in the images using the annotation tools available at MakeSenseAI (www.makesense.ai/). Bounding boxes are usually defined as the minimal-area rectangle that encompasses the whole object region (Murrugarra-Llerena *et al.*, 2022). Resultant annotations were saved in text (.txt) file format, containing coordinates and class labels for each annotated region. The annotation process was double-checked by a plant pathologist to ensure accuracy and consistency in identifying MYMV infected regions and distinguishing them from other anomalies.

To enhance the robustness and generalization of the object detection model, the given data augmentation techniques were applied to the original dataset: a) Horizontal Flip- The images were flipped horizontally, mirroring the content. This augmentation helps the model generalize to symmetrically occurring patterns, such as plant diseases appearing on either side of the leaves. b) Random brightness and contrast adjustment- Brightness and contrast of the images were randomly adjusted within a defined range. c) Rotation-Images were rotated randomly within a range of ± 15 degrees. This augmentation addresses the variability in the orientation of leaves and plants during data acquisition. d) Random Scaling Images were randomly scaled by a factor within $\pm 20\%$ of the original size to simulates variability in the distance between the camera and the plant during image capture. All augmentations were performed using the 'Albumentations' library in Python. Each transformation was applied individually to the original dataset, resulting in multiple augmented versions of each image. Bounding box annotations were adjusted dynamically to ensure their accuracy relative to the transformations applied. To ensure consistency and quality in the dataset, preprocessing steps were carried out on the augmented images and their corresponding annotations. All images were normalized by dividing each pixel value by 255. Images that were unreadable or corrupted were identified and excluded from the dataset.

As a result of the above process, a dataset containing 4480 images was generated having one class *i.e.* MYMV.

This dataset formed the basis for developing and assessing the automated disease identification system. It supplied the essential ground truth data required for deep learning algorithms to learn and generate accurate predictions.

YOLOv8

YOLOv8 (Jocher *et al.*, 2023) is a state-of-the-art object detection algorithm renowned for its efficiency and accuracy in real-time detection tasks. YOLO integrates Convolution Neural Networks (CNNs) as its backbone architecture, enabling it to efficiently process input images while preserving spatial relationships and feature representations crucial for accurate object detection. Building upon its predecessors, YOLOv8 incorporates advancements in network architecture and training techniques, enabling robust detection of objects across various domains, including agriculture, surveillance and autonomous driving. Its single-pass architecture facilitates rapid inference without compromising detection quality, making it a popular choice for applications requiring both speed and precision.

Experimentation

The Soybean MYMV dataset comprised 4480 bounding box annotated images depicting various stages of crop. This dataset was randomly split into training and validation sets, with 3584 images allocated for training and 896 reserved for validation. Utilizing a transfer learning approach, all variants of the YOLOv8 object detector were trained. The original images were resized to a spatial resolution of 640 x 640 pixels to align with the specifications of the YOLOv8 network architectures. The experimentation, conducted in Python 3.10.5, adhered to default hyperparameter settings as specified in the official implementations (<https://github.com/ultralytics/ultralytics>). Training of the YOLO networks was carried out on the Kaggle platform (www.kaggle.com) utilizing T4x2 as an accelerator, with a batch size of 32 images. Each network underwent training for 150 epochs with early stopping criteria, such that if there is no significant improvement over 30 epoch, the training process will stop.

Performance evaluation

The accuracy of an object detection model can be evaluated against metrics such as precision (P), recall (R), F1-score, average precision (AP) and mean average precision (mAP) (Padilla *et al.*, 2021). The confusion matrix in Table 1 classifies all test results into True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The P, R, F1 values and mAP are calculated as shown in Equations.

$$P = \frac{TP}{TP+FP}$$

$$R = \frac{TP}{TP+FN}$$

$$F1 \text{ Score} = \frac{2PR}{P+R}$$

The mAP metric, the most extensively used in object detection research, was used in this study, which is calculated using the following equations:

$$AP = \int_0^1 P(R) dR$$

$$mAP = 1/n \sum_{i=1}^n AP$$

Where,

n= Number of classes.

mAP= Indicates the average AP across three weed classes.

The more the area under the PR curve, the better the accuracy of object detection. The parameter corresponding to the threshold of bounding box area can be defined as mAP@0.5 that indicates mean average precision at IoU threshold of 50%. Similarly, mAP@0.5:0.95 indicates mean average precision over IoU threshold starting from 50% to 95%. Additionally, inference time and model parameters were considered to evaluate the computational efficiency and complexity of the models. Table 2 provides an overview of evaluation matrices used in the study.

RESULTS AND DISCUSSION

The data presented in Table 3. showcases the exemplary performance of all YOLOv8 variants on the MYMV disease detection dataset. Notably, YOLOv8x exhibited the highest precision at 98.1%, indicating its proficiency in minimizing false positives. Conversely, YOLOv8s achieved the highest recall of 78.6%, demonstrating its effectiveness in capturing a greater proportion of true positives. Mean Average Precision (mAP) emerged as a pivotal metric for evaluating object detection models, with YOLOv8s attaining the highest mAP (85.3%) at 0.5 Intersection over Union (IoU) threshold and YOLOv8m achieving the highest mAP (73.2) at 0.5-0.95 IoU threshold. Fig 2 represents mAP@0.5 curves of all models during whole training process. These results highlight the ability of YOLOv8s to accurately detect objects with high overlap, while YOLOv8m excels in achieving comprehensive detection across a range of IoU thresholds. Additionally, YOLOv8s demonstrated the highest F1-score (86.9), underscoring its balanced performance in terms of precision and recall.

However, in real-world applications, considerations extend beyond accuracy metrics to include computational efficiency. YOLOv8s emerged as a compelling choice, boasting the lower inference time of 5.6 milliseconds and the least number of model parameters, indicative of its reduced complexity. A faster inference time ensures that the model can be effectively used in time-sensitive scenarios. Optimizing inference time is essential for achieving efficiency and maintaining user engagement in practical deployments.

YOLOv8s also showcased a commendable mAP@0.5 of 97.2%, comparable to that of YOLOv8x. Hence, YOLOv8s emerges as a superior option for resource-constrained environments, offering a balance between accuracy and computational efficiency. Fig 3 depicts the images where MYMV disease is detected by YOLOv8s object detector.

These findings underscore the importance of considering both performance metrics and practical constraints when selecting models for real-world deployment. The results of the study are in line with the results achieved by Ahmed and Abd-Elkawy, (2024); Han *et al.* (2024), similarly with diagnosis of major foliar diseases in black gram (*Vigna mungo* L.) using Convolution Neural Network (Kalpana *et al.*, 2023), faba bean leaf disease (Jeong and Na, 2024) and wilting in soybean (Na and Na, 2024).

Future research thrusts could focus on enhancing the performance and applicability of YOLOv8-based disease detection models in soybean crops. Areas for exploration include optimizing model parameters for improved detection accuracy, integrating multi-disease detection capabilities to address diverse crop health challenges and conducting field validation studies to assess real-world performance under varying conditions. Additionally, investigating methods to streamline inference time and



Fig 1: Representative images of the soybean MYMV infection.

Table 1: Confusion matrix.

Actual class	Predicted class		
	Positive	Positive	Negative
Positive		True positive (TP)	False negative (FN)
Negative		False positive (FP)	True negative (TN)

Table 2: Performance evaluation matrices used in the study.

Precision (P)	Measures the accuracy of positive predictions, indicating the proportion of true positives among all predicted positives.
Recall (R)	Measures the ability to detect positive instances, representing the proportion of true positives identified correctly.
Mean average precision (mAP)	Computes the average precision across all classes, providing a comprehensive evaluation of detection performance.
F1-score	Harmonic mean of precision and recall, offering a balanced assessment of a model's performance.
Inference time	Time taken by the model to make predictions on new data, reflecting the computational efficiency of the detection process.
Model parameters	Number of parameters in the model, indicating its complexity and resource requirements during training and inference.

Table 3: Precision, recall, mAP@0.5, mAP@0.5-0.95, F1-Score, inference time and model parameter comparison of different weed detection models.

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5-0.95 (%)	F1-Score (%)	Inference time (ms)	Model Parameters (millions)	GFLOPs
YOLOv8l	95.1	76.7	81.9	69.8	84.9	16.3	43.6	164.8
YOLOv8m	97.1	75.9	84.4	73.2	85.2	11.9	25.8	78.7
YOLOv8n	93.5	76.5	84.1	66.30	84.2	3.6	3.1	8.1
YOLOv8s	97.2	78.6	85.3	72.2	86.9	5.6	11.1	28.4
YOLOv8x	98.1	76.5	83.6	73	86.0	24.2	68.1	256.2

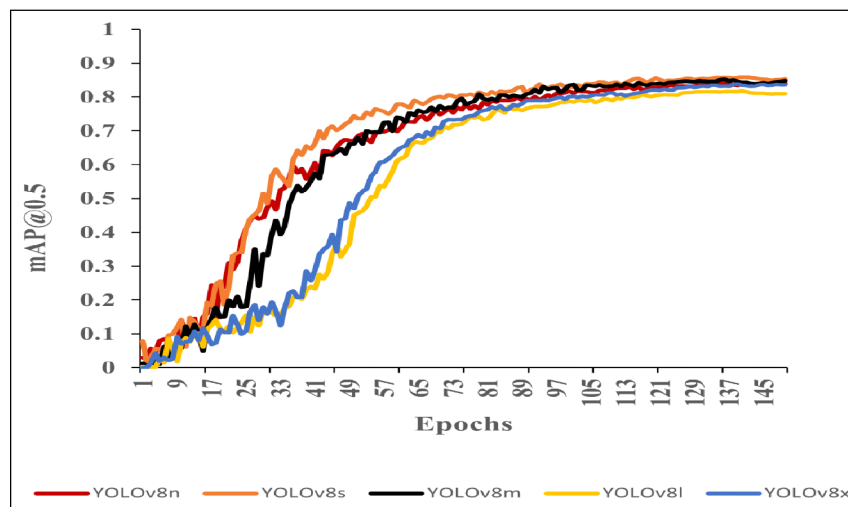


Fig 2: Training curves of mAP@0.5 YOLOv8 models.

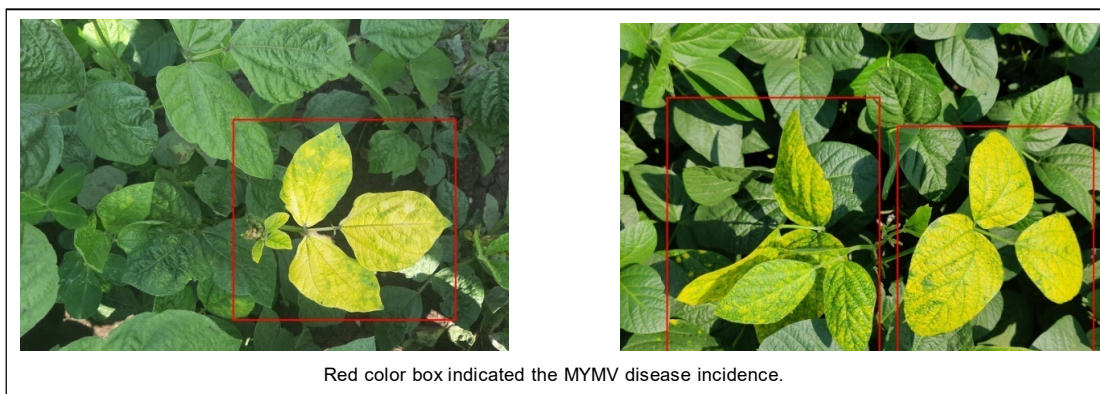


Fig 3: Representative images detected using YOLOv8n object detector.

model size without compromising accuracy could facilitate real-time deployment in practical agricultural settings. Incorporating multispectral and hyperspectral imagery could further improve the detection of disease at early-stage. Collaborative research initiatives involving agronomists and plant pathologists could contribute to the development of holistic solutions that leverage both advanced technologies and traditional agricultural practices, ultimately advancing sustainable crop production and food security.

CONCLUSION

The present study investigated the performance of various YOLOv8 variants for the detection of Yellow Mosaic Disease caused by Mungbean Yellow Mosaic Virus in soybean crop. Leveraging our newly created dataset, which contains 4480 annotated images depicting different stages of crop growth, we conducted a comprehensive evaluation of the YOLOv8 models. Our findings showcase the exemplary performance of all YOLOv8 models on the MYMV disease detection dataset, with each variant demonstrating strength in different aspects of detection accuracy and computational efficiency. Notably,

YOLOv8x exhibited outstanding precision, while YOLOv8s achieved remarkable recall rates. Mean Average Precision (mAP) emerged as a critical metric, with YOLOv8s excelling in accurately detecting objects with high overlap and YOLOv8m showcasing comprehensive detection across different Intersection over Union (IoU) thresholds (mAP@0.5-0.95). Additionally, YOLOv8s emerged as a compelling choice for real-world deployment due to its minimal inference time and model complexity, offering a balance between accuracy and computational efficiency. These findings underscore the importance of considering both performance metrics and practical constraints when selecting models for disease detection in agricultural settings.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

- Ahmed, R. and Abd-Elkawy, E.H. (2024). Improved tomato disease detection with YOLOv5 and YOLOv8. *Eng. Technol. Applied Scientific Research*. **14**: 13922-13928.

- Bazzi, A., Slock, D.T. and Meilhac, L. (2017). A Newton-type Forward Backward Greedy Method for Multi-snapshot Compressed Sensing. In 2017 51st Asilomar Conference on Signals, Systems and Computers IEEE. (pp. 1178-1182).
- Bazzi, A., Slock, D.T., Meilhac, L. and Panneerselvam, S. (2016). A Comparative Study of Sparse Recovery and Compressed Sensing Algorithms with Application to AoA Estimation. In 2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE. (pp. 1-5).
- Han, R., Shu, L. and Li, K. (2024). A method for plant disease enhanced detection based on improved YOLOv8. In: 2024 IEEE 33rd International Symposium on Industrial Electronics (ISIE), Ulsan, Korea. pp. 1-6.
- Heltin Genitha, C., Dhinesh, E. and Jagan, A. (2019). Detection of Leaf Disease using Principal Component Analysis and Linear Support Vector Machine. In: 2019 International Conference on Advanced Computing (ICoAC). pp. 1-5.
- Jeong, H.Y. and Na, I.S. (2024). Efficient faba bean leaf disease identification through smart detection using deep convolutional neural networks. *Legume Research*. **47(8)**: 1404-1411. doi: 10.18805/LRF-798.
- Jocher, G., Chaurasia, A. and Qiu, J. (2023). YOLO by Ultralytics. Kalpana, M., Karthiba, L., Senguttuvan, K. and Parimalarangan, R. (2023). Diagnosis of major foliar diseases in black gram (*Vigna mungo* L.) using convolution neural network (CNN). *Legume Research*. **46(7)**: 940-945. doi: 10.18805/LR-5083.
- Mandhare, V.K. and Gawade, S.B. (2010). Effect of seed-borne soybean mosaic virus infection on quality and yield parameters in soybean. *Legume Research*. **33**: 43-49.
- Mathew, M.P. and Mahesh, T.Y. (2022). Leaf-based disease detection in bell pepper plant using YOLO v5. *Signal Image Video Process*. **16**: 841-847.
- Murrugarra-Llerena, J., Kirsten, L. N., and Jung, C. R. (2022). Can We Trust Bounding Box Annotations for Object Detection? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4813-4822).
- Na, M.H. and Na, I.S. (2024). Detection and classification of wilting in soybean crop using cutting-edge deep learning techniques. *Legume Research*. **47(10)**: 1723-1729. doi: 10.18805/LRF-797.
- Nasution, S.W. and Kartika, K. (2022). Eggplant disease detection using YOLO algorithm with telegram notification. *International Journal of Engineering, Science and Information Technology*. **2(4)**: 127-132.
- Padilla, R., Netto, S.L. and da Silva, E.A.B. (2020). A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil. pp. 237-242.
- Rohit, M.K., Gurupad, B.B., Lokesh, B.K. and Suma, M. (2023). Prevalence of mungbean yellow mosaic virus on greengram in northern parts of Karnataka. *Journal of Farm Sciences*. **36(1)**: 47-49.
- Sangaiah, A.K., Yu, F.N., Lin, Y.B., Shen, W.C. and Sharma, A. (2024). UAV T-YOLO-Rice: An enhanced tiny YOLO network for rice leaf disease detection in paddy agronomy. *IEEE Trans. Network Science Engineering*. pp 1-16.
- Shill, A. and Rahman, M.A. (2021). Plant disease detection based on YOLOv3 and YOLOv4. In: 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI). pp. 1-6.
- Singh, V.B., Haq, Q.M.R. and Malathi, V.G. (2013). Antisense RNA approach targeting Rep gene of Mungbean yellow mosaic India virus to develop resistance in soybean. *Archives of Phytopathology and Plant Protection*. **46**: 2191-2207.
- Soeb, Md.J.A., Jubayer, Md.F., Tarin, T.A., Al Mamun, M.R., Ruhad, F.M., Parven, A., Mubarak, N.M., Karri, S.L. and Meftaul, I.Md. (2023). Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Scientific Reports*. **13**: 6078.
- Uoc, N.Q., Duong, N.T. and Thanh, B.D. (2022). A novel automatic detecting system for cucumber disease based on the convolution neural network algorithm. *GMSARN International Journal*. **16**: 295-302.
- Yang, S., Yao, J. and Teng, G. (2024). Corn leaf spot disease recognition based on improved YOLOv8. *Agriculture*. **14**: 666.
- Zhao, Y., Gong, L., Huang, Y. and Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. *Computers and Electronics in Agriculture*. **127**: 311-323.