



RESEARCH ARTICLE

Comparative evaluation of YOLO algorithms for detection and classification of rice leaf diseases using field-collected datasets

Ashikur Rahman¹, Chung Gwo Chin^{1*}, Ershadul Haque² & Lee It Ee¹

¹Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya 63100, Selangor, Malaysia

²School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, New South Wales 2795, Australia

*Correspondence email - gcchung@mmu.edu.my

Received: 21 October 2025; Accepted: 29 April 2026; Available online: Version 1.0: 02 July 2026

Cite this article: Ashikur R, Chung GC, Ershadul H, Lee IE. Comparative evaluation of YOLO algorithms for detection and classification of rice leaf diseases using field-collected datasets. Plant Science Today (Early Access). <https://doi.org/10.14719/pst.12376>

Abstract

Rice is the staple food for more than half of the global population. However, the productivity and quality of rice or grain dropped significantly due to leaf diseases. These diseases are difficult to identify through manual processes, which are time-consuming, labour-intensive and often inaccurate, particularly in rural farming communities. With recent advances in computer vision technology, object detection algorithms, namely the you only look once (YOLO) family, can provide high-speed, high-accuracy solutions for automated plant disease detection. This study evaluates four YOLO variants, such as YOLOv5, YOLOv7, YOLOv8 and YOLOv11, using about 1500 field images collected in Bangladesh. The data represent four major rice leaf diseases, such as bacterial leaf blight, brown spot, leaf blast and sheath blight. Data pre-processing, including image annotation and data augmentation, was conducted before model training and was followed by the training of the YOLO models. All the models were trained with the same hyperparameters and their performance was evaluated using standard metrics, such as F1 scores, precision, recall and mean average precision (mAP). According to experimental findings, the YOLOv7 recorded the highest performance based on F1 score of 0.77 and mAP of 0.85 in comparison with the rest of the variants. The results suggest that YOLOv7 will be the most appropriate to use instead of other models in the detection of rice leaf disease in real-time, which can be utilised in precision agriculture and mobile-based disease management systems.

Keywords: image classification; machine learning; rice leaf disease detection; YOLO

Introduction

Rice cultivation is recognised as one of the most significant agricultural activities on the planet because of its commercial value and cultural importance as a staple food crop. More than 2.7 billion people worldwide rely on rice as their primary source of food. According to the predictions of the UN Food and Agriculture Organization (FAO), global rice output is expected to reach a record 556.4 million tonnes (t) between 2025 and 2026 (1). Bangladesh ranks as the world's third-largest producer and approximately 169 million people depend on rice as their primary staple food (2). Rice production has shown a steady increase over recent years. Statistics indicate that the production of rice is increasing day by day. Therefore, research related to rice is becoming a vibrant field for the food industry. Another significant challenge to rice production is the emergence of diseases affecting rice leaves. The yield of rice has decreased considerably due to diseases caused by bacteria, viruses or fungi that impact the rice leaves. These organisms cause immense destruction to the crop and significantly lower its yield. Some of the rice leaf diseases with capacities to compromise such foliage consist of bacterial leaf blight, brown spot, leaf blast, leaf scald, leaf tungro, leaf ufra, narrow brown leaf spot and sheath blight, among others (3). This study focuses on four major diseases: bacterial leaf blight, brown spot, leaf blast and sheath blight. These types of rice leaf diseases are typical of Bangladesh. Bacterial leaf blight is caused by

Xanthomonas oryzae pv. *oryzae*, while brown spot and leaf blast are caused by fungal pathogens. Sheath blight, caused by *Rhizoctonia solani*, is also a major disease affecting rice production (4–9).

Therefore, computer vision is a more adequate method for disease detection and classification. The technology of convolutional neural networks (CNN) was initially introduced with this purpose and it is a part of the larger class of artificial neural networks (10). Transfer learning improves model performance when large datasets are available (11–13). On the other hand, there are substantial versions of the you only look once (YOLO) family, such as YOLOv5, YOLOv8 and YOLOv9, that have been developed and trained on large-scale real-world image datasets, enabling improved accuracy, speed and generalisation performance across diverse object detection tasks (14).

You only look once algorithms have proven to be very fast and accurate when it comes to object detection in real time. The current study outlines a comparative performance evaluation of several advanced YOLO variants, including YOLOv5, YOLOv7, YOLOv8 and YOLOv11. It also stresses their ability to identify and classify illnesses of rice leaves. Despite the availability of several deep learning (DL) models, comparative evaluation of recent YOLO variants using real field datasets remains limited. Therefore, this study compares YOLOv5, YOLOv7, YOLOv8 and YOLOv11 for the

detection of rice leaf disease using field-collected images. The objective is to identify the most accurate and reliable model for real-time disease detection.

A wide range of methodologies has been explored for classification and detection of rice leaf diseases based on machine learning (ML) and deep learning (DL) models. According to the findings of previous studies, there are nine major rice leaf diseases worldwide, such as bacterial leaf blight, brown spot, leaf blast, leaf scald, leaf smut, leaf tungro, leaf ufra, narrow brown leaf spot and sheath blight. However, a field investigation in Bangladesh revealed that the four most prevalent diseases, including bacterial leaf blight, brown spot, leaf blast and sheath blight, are in practical agricultural environments. (3).

With the advancement of artificial intelligence (AI) and computer vision techniques, they became dominant for the automated diagnosis of rice leaf diseases. Convolutional neural network architectures such as Inception and ResNet have been used for rice disease classification, applying transfer learning approaches that significantly improved classification accuracy while decreasing the requirement for large-scale training from scratch (15). For instance, in one study, the University of Oxford developed the VGG-16 CNN model, which was trained and evaluated using datasets collected in paddy fields and web sources, achieving an accuracy of 92.42 % (16). In another study, three diseases including bacterial leaf blight, brown spot and leaf blast, were investigated by using a corpus of 1649 leaf images and a distributed event-learning model was proposed that suggested 98 % higher accuracy (17). In addition, a system based on the internet of things (IoT) and edge intelligence has been suggested for real-time disease detection (18). The system was tested on the 300 rice leaf images covering three disease types, such as brown spot, rice hispa and leaf blast and obtained an accuracy of 97.50 %.

Early studies focused on the conventional ML methods, such as support vector machines (SVM), Naïve Bayes and k-nearest neighbour. They were widely applied for classification and feature

extraction tasks. For example, SVM was employed using texture, colour and shape features obtained as segmented leaf regions, reaching an accuracy of 97.2 % for bacterial leaf blight, sheath and leaf blast detection (19). Comparative studies have shown that ML techniques can achieve reasonable classification accuracy when conditions are controlled, but tend to be sensitive to changes in illumination, background noise and strategies of feature selection (20).

Recently, the YOLO family of object detection models has been extensively utilised in the detection of agricultural diseases. For example, a YOLO-based method utilised on 200 datasets reported that it achieved about 90 % accuracy (21). Moreover, an herb species recognition neural network model, which was built on the basic YOLO, was used to classify four herb species, such as Mehndi, betel leaf, mint and aloe vera leaves and obtained an approximately 95 % accuracy (22). Recent studies have demonstrated improved disease detection accuracy using advanced YOLO variants such as YOLOv5, YOLOv7 and YOLOv8 (23–32). However, comparative evaluation of recent YOLO variants using real field datasets remains limited. Most previous studies focused on single models or laboratory datasets. Therefore, benchmarking multiple YOLO models using field-collected rice leaf images is necessary. The performance of recent models in the literature are summarised in Table 1.

Although the current research on developing DL models in rice leaf disease detection shows several promising signs, there are a number of key limitations that can be identified. To begin with, the absence of fair benchmarking is evident and few studies conduct comparative analyses of several recent YOLO architectures, including YOLOv5, YOLOv7 and YOLOv8, with the same experimental conditions. Second, validation in the field is not adequately developed, as most of the datasets are obtained in the laboratory, which limits the ability of trained models to generalise to real farm conditions. Third, cross-study comparisons are inaccurate and potentially misleading due to discrepancies in assessment protocols, among other things, the size of the datasets, methods of pre-processing and the hyperparameters, performance measures,

Table 1. Summary of related studies on plant and rice leaf disease detection using machine learning models

Reference	Proposed method	Dataset	Diseases identified	Accuracy (%)	Additional notes
(14)	YOLO (v5, v8 and v9)	8858 images	22 disease types related to apples, cucumbers, tomatoes and grapes	50.65, 91.81 and 93.1	Outperformed YOLOv9
(15)	Inception ResNet v2	5200 images	Leaf blast, bacterial blight, brown spot	95.67	Transfer learning strategy
(16)	VGG16	1649 images	Brown spot, leaf blast, bacterial leaf	92.46	Dataset compiled from field samples and online
(20)	YOLO	200 images	Leaf blast, brown spot	90 (Leaf blast), 70 (Brown spot)	Custom dataset
(17)	Lightweight AI	300 images	Hispa, brown spot, leaf blast	97.5	-
(18)	SVM	-	Bacterial leaf blight, sheath blight	97.2	Image processing techniques used
(22)	YOLOv5	86 images	-	96.5	Training across five different epochs
(26)	YOLOv5	2050 images	Mold detection	-	Outperformed YOLOv3 and YOLOv4
(28)	YOLOv5	5740 images	Black rot, fish scale disease, snow apple	91.93	Training on Plant Village dataset
(19)	SVM, Naïve Bayes and CNN	3355 images	Leaf blast, brown spot, healthy leaf	97.9	-
(29)	YOLOv4	Publicly available dataset and local dataset	Various crop diseases	99.21 (Public), 97.89 (Local)	Designed for crop disease recognition and identification
(30)	YOLOv8s	3622 original datasets of wheat disease photos	Multiple wheat diseases	93.28 with mAP of 0.5	New C2f-DCN module and the SCNet attention mechanism
(31)	MobileNetCA-YOLOv7	Dataset of 3773 rice pest and disease images	Rice pest and disease recognition	92.3 with mAP of 0.5	Lightweight network MobileNetV3 for feature extraction, reducing parameterization and incorporates the coordinate attention mechanism (CA)
(32)	YOLOv7, YOLOv8 and YOLO-NAS	1000 images of rice plant leaf diseases	ALS, CBB, CBMV, bean rust, anthracnose	97.9	Deployed YOLO-NAS annotation models into an Android app, validating their effectiveness.

etc. Finally, there has been a lack of focus on robustness analysis, especially in determining the ability of the different variants of YOLO to act on real-world environments as related to changing illumination and occlusion, the complexity of the surrounding environment and the natural environmental variability. All these gaps point to the importance of a unified, field-based and holistic benchmarking study that would guarantee the validity of the performance measurement and viability of the practical implementation. To overcome these gaps, this paper conducts a systematic evaluation of the four recently developed YOLO architectures, including YOLOv5, YOLOv7, YOLOv8 and YOLOv11, on a dataset of field-collected rice leaf disease to benchmark them. It is because all the models are trained and tested in the same experimental conditions so that there is a fair comparison. The research gives a detailed performance analysis based on various detection measures, such as precision, recall, F1-score and average precision mean. This work will bring credible findings on the comparative performance of the latest YOLO architectures and the most appropriate model to use in real-life applications of rice disease detection.

Materials and Methods

This section outlines the methodological framework adopted for the detection and classification of rice leaf diseases using shallow-trained YOLO models. A comprehensive pipeline that includes data collection, pre-processing, model training and evaluation are depicted in Fig. 1.

Study area and data collection

The dataset was collected by taking the photographs of the rice leaves in Barahipur, Feni, Bangladesh (approximate coordinates: 23.025 °N latitude and 91.395 °E longitude) as shown in Fig. 2. Data collection was conducted during the aman rice-growing season from June–November, which coincides with the monsoon period

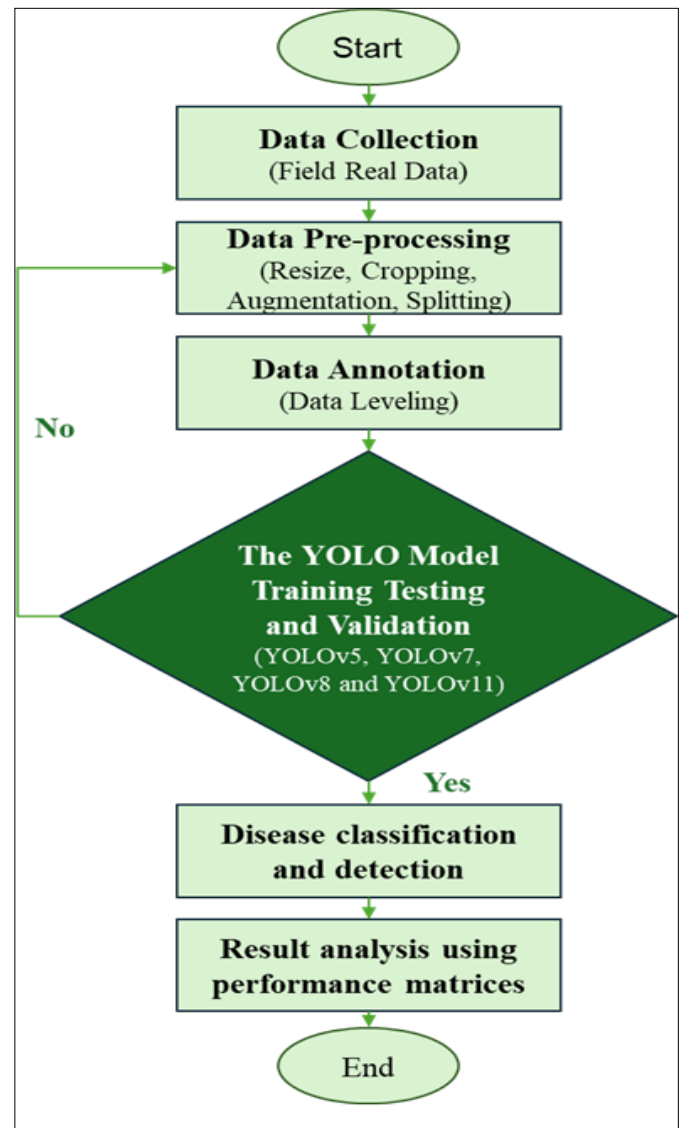


Fig. 1. Flow chart of the proposed approach.

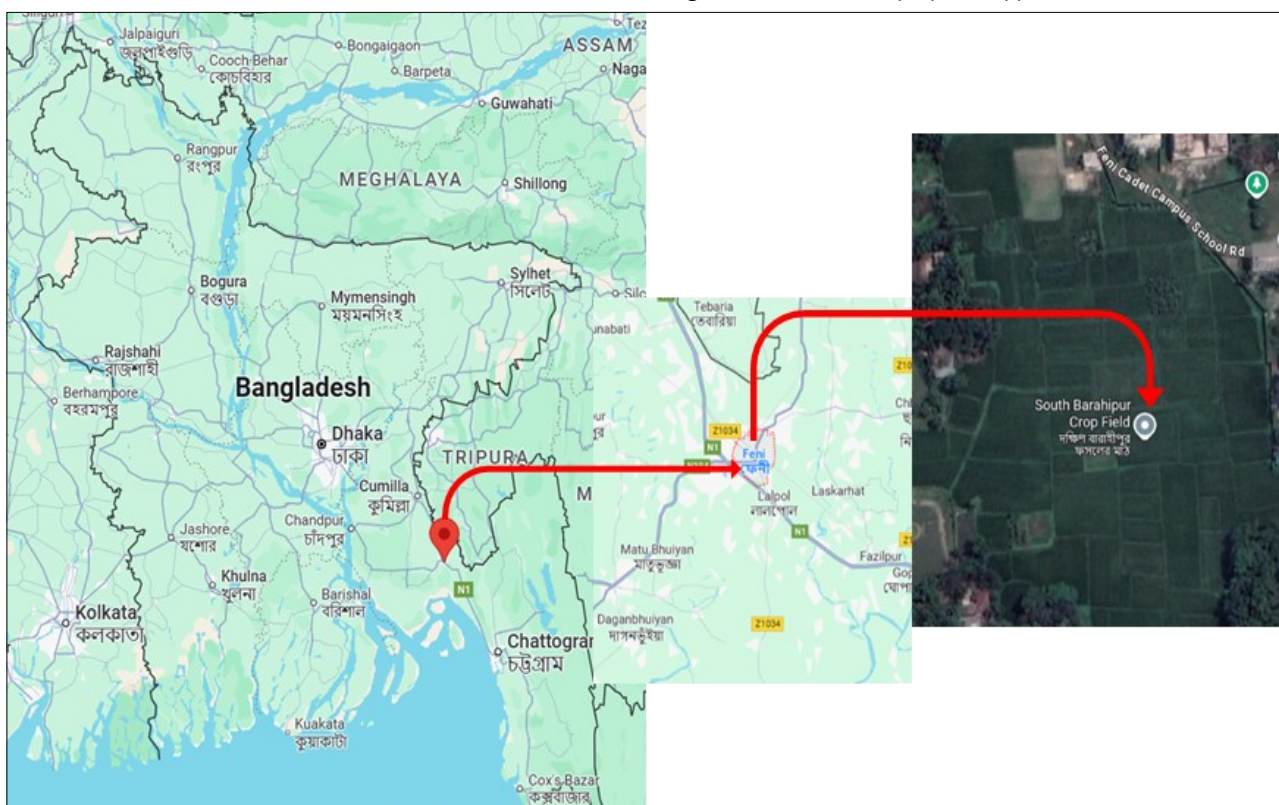


Fig. 2. Location of collected data.

and provides favourable environmental conditions for the development of common rice leaf diseases. Images were captured using a Redmi Note 10 smartphone equipped with a 48-megapixel camera. The raw dataset consisted of approximately 1500 rice leaf images in total and was distributed into two primary categories: healthy and infected samples. The infected class included four major diseases, such as bacterial leaf blight, brown spot, leaf blast and sheath blight, as illustrated in Fig. 3.

Data preparation

All images were resized to 416×416 pixels before training. Standardising the image dimensions helps maintain consistent feature scaling, reduces computational complexity and improves training stability. The impact of this resizing procedure is shown in Fig. 4. Subsequently, the dataset was divided into three categories: 70 % training set, 20 % validation set and 10 % testing set, that ensures adequate data for model learning, hyperparameter tuning and unbiased performance evaluation.

Annotation and augmentation

Images were annotated using the Roboflow platform (Fig.5). The images were opened one at a time. Then, the boundaries of the target lesion were manually drawn using a rectangle to identify the location of the target lesion using the YOLO. It requires class label, x-center, y-center, width and height. The labels attributed to each occurrence were healthy, brown spot, leaf blast, sheath blight and bacterial leaf blight, among others. The formatted data had annotations that were exported in the YOLO text format (.txt). Data augmentation was carried out with the help of Roboflow to improve the diversity of data and reduce overfitting caused by the small size of the sample. Several augmentation methods, including horizontal and vertical flipping, random cropping, rotation (90° and 180°), brightness adjustment, contrast enhancement and colour jittering, were used (Fig. 6). These changes produced diverse training samples and maintained semantics within classes, hence enhancing the strength and the overall generalisation power of the trained detection models.

Model architecture and training

In this study, four YOLO architectures, including YOLOv5, YOLOv7, YOLOv8 and YOLOv11, were trained and evaluated under identical conditions to ensure a fair comparison. All models were initialised with pre-trained common objects in context (COCO) weights and subsequently fine-tuned on the rice leaf disease dataset. The input images were resized to 416×416 pixels and a batch size of 16 was used during training. Each model was trained for 100 epochs with an initial learning rate of 0.001, using the adaptive moment estimation (Adam) optimiser and an early stopping strategy with a patience of 20 epochs to reduce overfitting. Training was conducted on an NVIDIA RTX 3060 GPU with 12 GB of VRAM. Even though traditionally YOLO models are trained through the stochastic gradient descent (SGD) with momentum, this paper used the Adam optimiser because of the relatively small size of the dataset and faster convergence required during fine-tuning. The mechanism of adaptive learning rate in Adam arranges dynamically changing the parameter updates according to the first and second-order moment approximations and has the potential to stabilise training and enhance convergence rate in the low-data conditions. The comparative assessment is also fair and consistent, as every four architectures were trained with the same optimiser and hyperparameters.

Architecturally, YOLOv5 serves as a baseline, using a cross-stage partial (CSP) backbone with a PANet neck and a unified detection head (Fig. 7). Model YOLOv7 expands on this backbone by introducing the extended efficient layer aggregation network (E-ELAN), which improves gradient flow and feature reuse, resulting in faster convergence and higher detection accuracy. Model YOLOv8 adopts a decoupled head structure that separately processes classification and localisation tasks, thereby improving performance in detecting small or overlapping lesions on rice leaves. The most recent iteration, YOLOv11, incorporates additional architectural changes targeted at improving the precision recall trade-off; however, its effectiveness in agricultural disease detection is still being investigated. By evaluating these four architectures within a standardised training framework, this study establishes a reliable benchmark for selecting the most effective YOLO variant for rice leaf disease detection in real-world agricultural conditions.

Results and Discussion

Performance metrics

The performance of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models was evaluated using standard object detection metrics. The analysis is based on traditional performance measures, including precision, recall, F1 score, accuracy, average precision (AP) and mean average precision (mAP). These measures make sure that the performance of the models is evaluated not just on the classification accuracy but also on the methods of localisation and strength in the real field conditions.

Intersection over union

Object detection performance depends on predicted bounding boxes overlap with ground truth annotations. This overlap is quantified using intersection over union (IoU):

$$\text{IoU} = \frac{\text{Area}(B_{\text{pred}} \cap B_{\text{gt}})}{\text{Area}(B_{\text{pred}} \cup B_{\text{gt}})} \quad (\text{Eqn. 1})$$

where, B_{pred} is predicted bounding box and B_{gt} is ground-truth bounding box. A prediction is considered a true positive (TP) if $\text{IoU} \geq \text{threshold}$ typical 0.5 and corrected background prediction is defined as true negative. Otherwise, it is classified as false positive (FP) and undetected objects are counted as false negative (FN).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (\text{Eqn. 2})$$

Precision measures the reliability of positive detections. High precision indicates low false-positive rates.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{Eqn. 3})$$

Recall evaluates the model's ability to detect all relevant disease instances.

$$\text{F1} = \frac{2 \times \text{Precision/Recall}}{\text{Precision} + \text{Recall}} \quad (\text{Eqn. 4})$$

F1-score balances both false positives and false negatives, providing a unified performance indicator.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (\text{Eqn. 5})$$

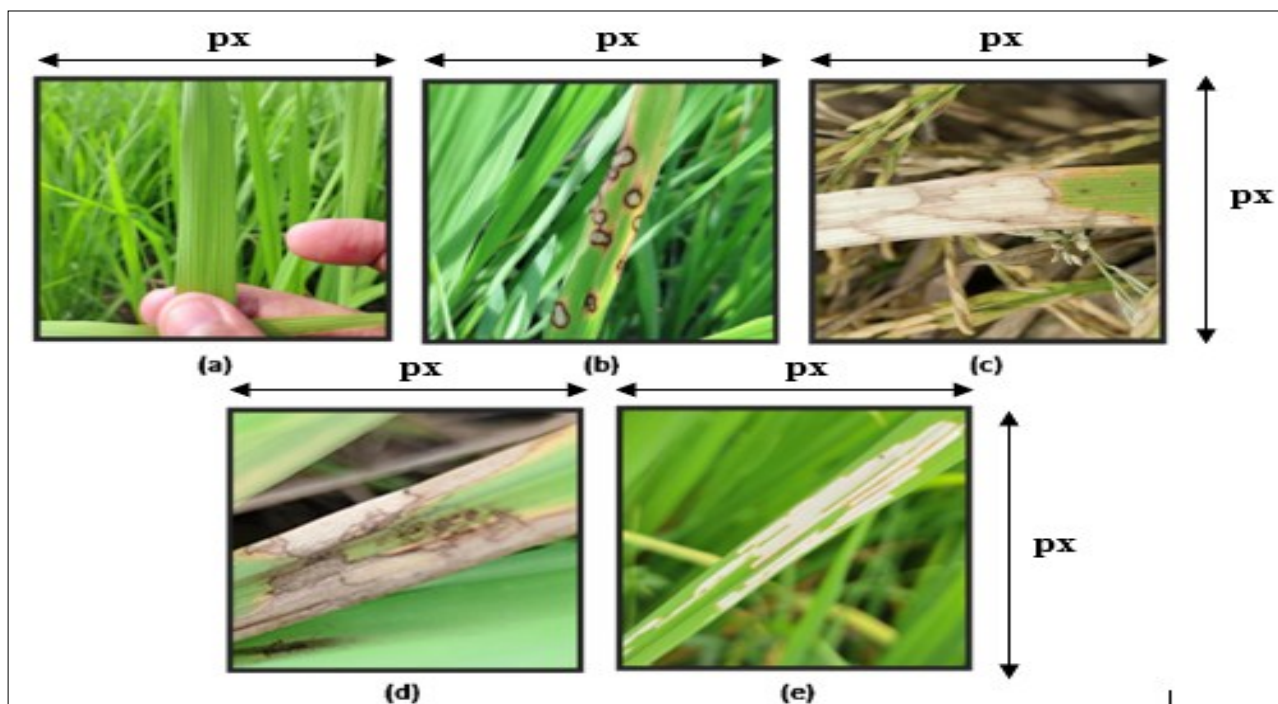


Fig. 3. (a) Healthy (b) brown spot (c) sheath blight (d) leaf blast and (e) bacterial leaf blight.

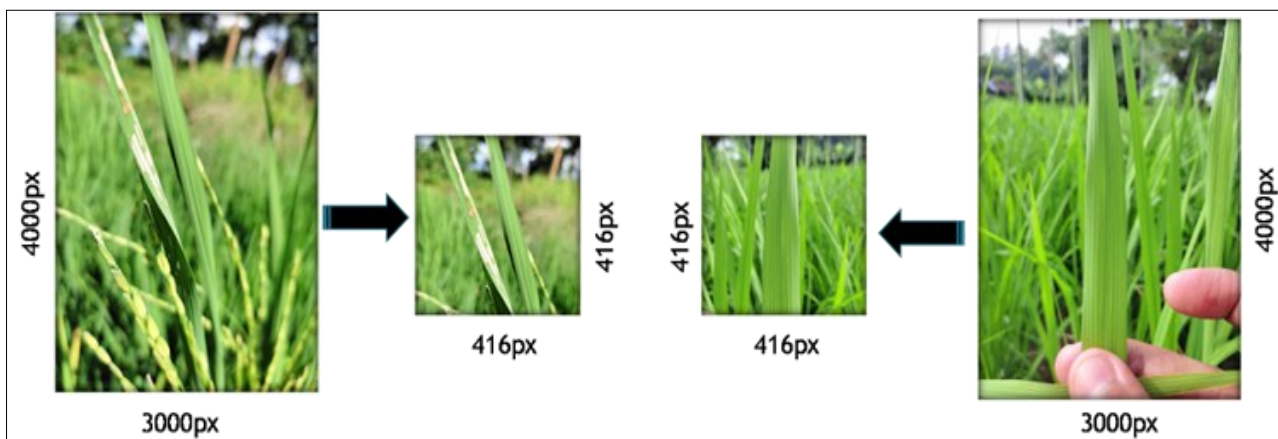


Fig. 4. Data resize and pre-processing.

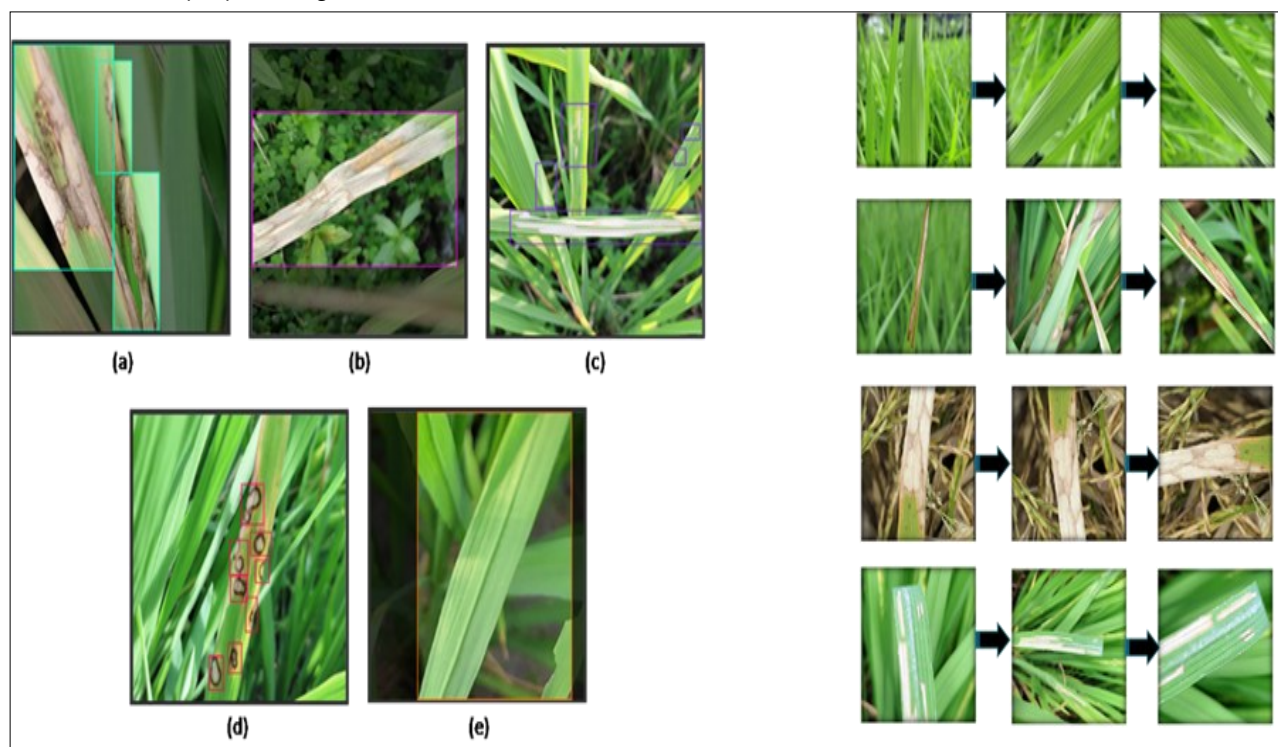


Fig. 5. Data annotation and augmentation.

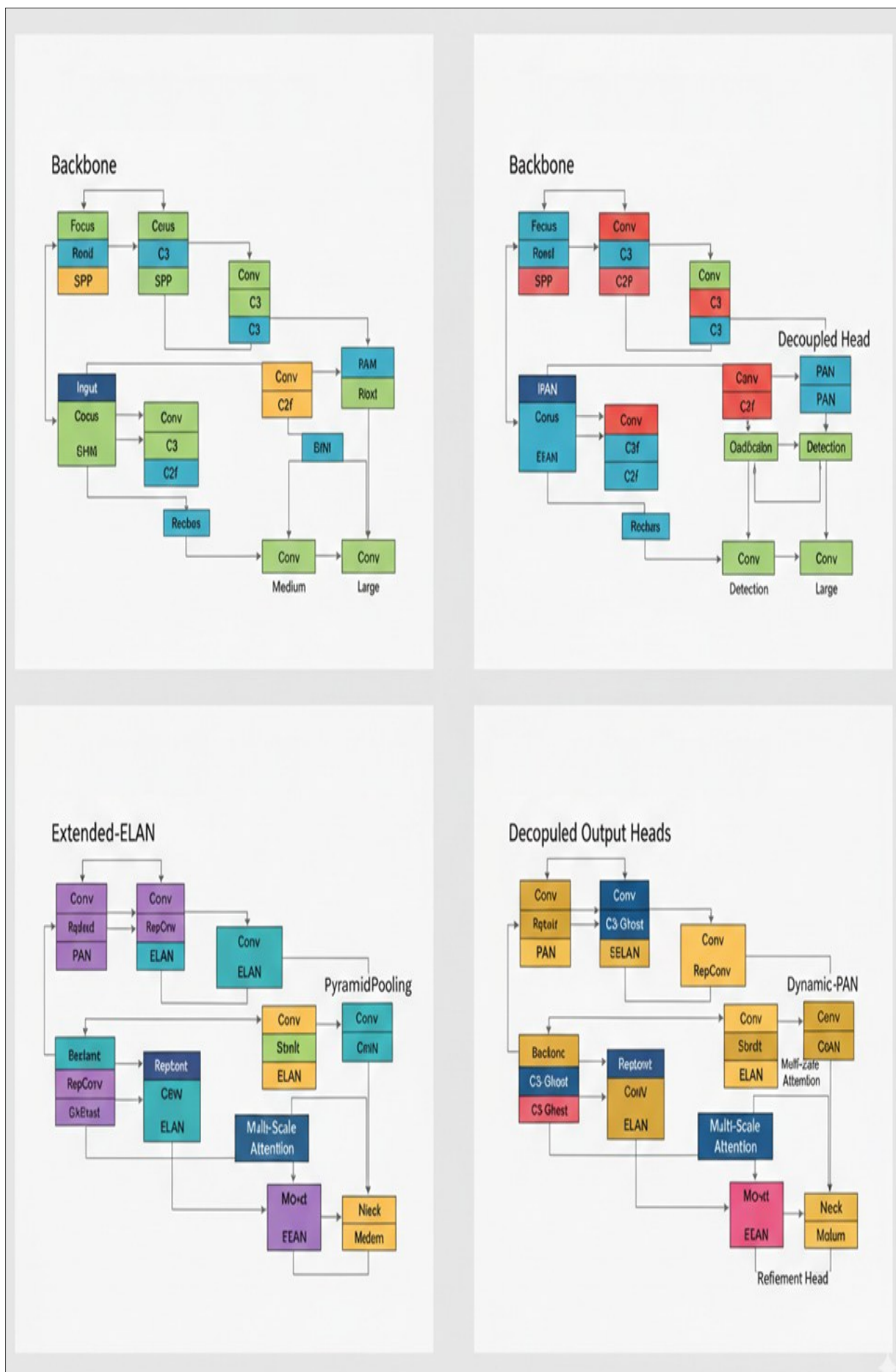


Fig. 6. You only look once (YOLO) model architecture.

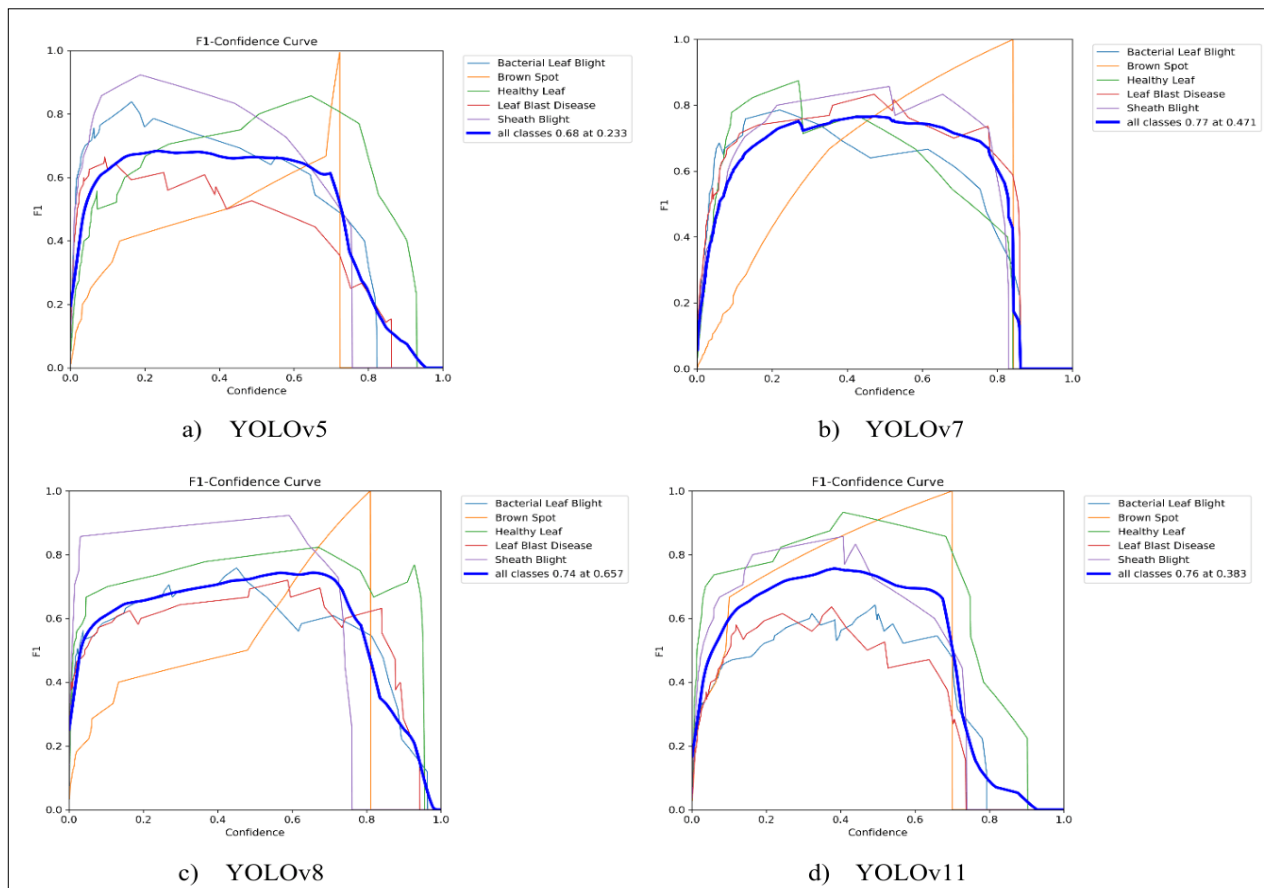


Fig. 7. F1 performance of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models.

Although widely used, accuracy alone is insufficient for object detection due to class imbalance and localisation sensitivity.

Average precision and mean average precision

In object detection, performance is more rigorously evaluated using AP, which illustrates the area under the precision-recall (PR) curve for each class:

$$AP = \int_0^1 \text{Precision}(\text{Recall}) d\text{Recall} \quad (\text{Eqn. 6})$$

The mAP evaluates overall detection performance by calculating the AP across all N classes:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (\text{Eqn. 7})$$

Two commonly used assessing standards are mAP at 0.5 and mAP at 0.5:0.95. The mAP at 0.5 metric computes mAP using a fixed IoU threshold of 0.5, whereas mAP at 0.5:0.95 averages the AP across multiple IoU thresholds ranging from 0.5–0.095 with a step size of 0.05. At a more permissive level, mAP at 0.5 is used where the emphasis being made on the detection ability. Whereas at a more rigorous level, mAP at 0.5:0.95 offers a more wholesome assessment of localisation accuracy. In this paper, mAP 0.5 was mostly chosen to match the literature on agricultural detection, but mAP at 0.5:0.95 was also cited to give a more detailed perspective on bounding box regression performance and resilience in real-field conditions. The matrices report an overall assessment of the classification and localisation error of each model.

Confusion matrices

The confusion matrices for YOLOv5, YOLOv7, YOLOv8 and YOLOv11 on the test dataset of rice leaf disease is shown in Fig. 8. In each, the matrices demonstrate the distribution of samples classified correctly and incorrectly in five categories: bacterial leaf blight, brown spot, healthy leaf, leaf blast and sheath blight. The values that are found to be correct on the diagonal denote the correctly identified cases, whereas the values that are found to be off diagonal denote misclassification. The findings indicate that YOLOv7 possesses the highest level of consistency and its diagonal dominance is strong in all the classes, meaning it has an equal level of precision and recall. It has a smaller number of misclassifications among visual-related diseases like brown spot and leaf blast. Model YOLOv8 is also effective and is confused with sheath blight and brown spot, which implies it is sensitive to the variation of the background. Model YOLOv11 has competitive accuracy on most classes and has a weaker generalisation than that of YOLOv7. Instead, YOLOv5 produces more false positives and cross-class confusion, particularly between the categories of healthy and diseased leaves, despite having the highest overall recall. The confusion matrices support the quantitative performance metrics, confirming that YOLOv7 is the most robust and reliable classification across all categories of rice leaf diseases, while YOLOv5 tends to overpredict disease presence and YOLOv8 and YOLOv11 are intermediately robust.

Confidence and optimisation insights

The F1 score curves for all four YOLO models under evaluation are depicted in Fig. 8. The correctly assessed YOLOv7 will have

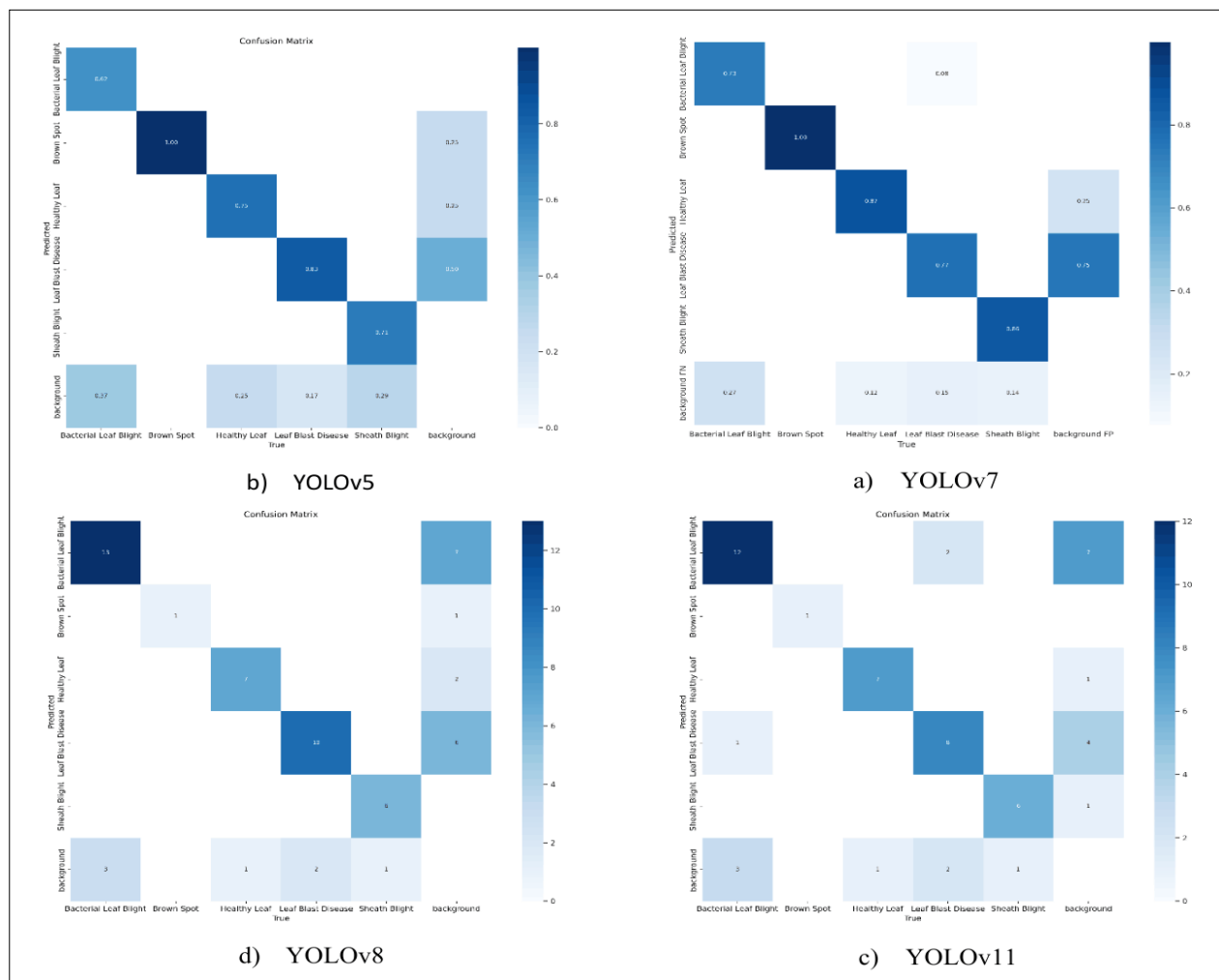


Fig. 8. Confusion matrices of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models for rice leaf disease detection.

the maximum F1 of 0.76 with the confidence threshold of 0.474, which is an ideal compromise between precision and recall. This confidence level will optimise the capability of the model to accurately characterise rice leaf diseases with the fewest inaccurate positive results and negative outcomes. Model YOLOv7 maintains higher precision across a range of confidence thresholds compared to other models (Fig. 9). Likewise, as demonstrated in Fig. 10, YOLOv7 has the best recall rate, detecting more true positives than YOLOv5, YOLOv8 and YOLOv11. This stability in recall reduces the hazard of undetected diseased samples.

Precision-recall analysis

The precision-recall (PR) curves of all four prefixes of the YOLO models under consideration are graphically shown in Fig. 11. The PR curve is an all-inclusive measure of the performance of this model in terms of classification across the different threshold settings. Model YOLOv7 is the only model that shows a high precision and recall ratio irrespective of the type of rice leaf disease affected in several categories. This shows that it is effective both in identifying positive cases correctly as well as reducing false negatives even with varied confidence thresholds. In contrast, YOLOv5 and YOLOv8 show significant recall decreases for specific disease classes. This decline is especially noticed in the cases of image noise, background clutter or different lighting conditions, which are the elements that are regular in the real-world agricultural conditions. These discrepancies indicate that, on one hand, YOLOv5, YOLOv8 and YOLOv11 can be highly precise; still, their ability to identify all the

positive cases is relatively lower. Owing to the above, the PR curve further suggests the overall better generalisation ability of YOLOv7 and strength under field conditions.

Subjective evaluation

The qualitative performance of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models in the training and testing stages using visual samples of the dataset are shown in Fig. 12 and 13. During the training step, the YOLOv7 model generates good-quality and continuous bounding boxes, as well as the proper class label of the observed rice leaf diseases (12). The model provides clear disease-affected areas with fewer false positives and misclassifications. This indicates strong learning capability and effective feature extraction during training. During the testing phase, YOLOv7 maintains its performance consistency, accurately detecting and classifying unseen images (Fig. 13). The bounding boxes are well-aligned with disease symptoms and class predictions show high reliability. In contrast, YOLOv5 and YOLOv8 occasionally exhibit bounding box misalignments or missed detections, especially in images with occlusions, variable lighting, or background noise. Model YOLOv11, while performing detection comparatively well, does not consistently surpass YOLOv7 in precision or clarity. These visual assessments reinforce the quantitative results by confirming that YOLOv7 is not only statistically superior but also visually more reliable in practical detection tasks.

The training and validation loss curves, as well as evaluation metrics (precision, recall, mAP) of YOLOv5, YOLOv7,

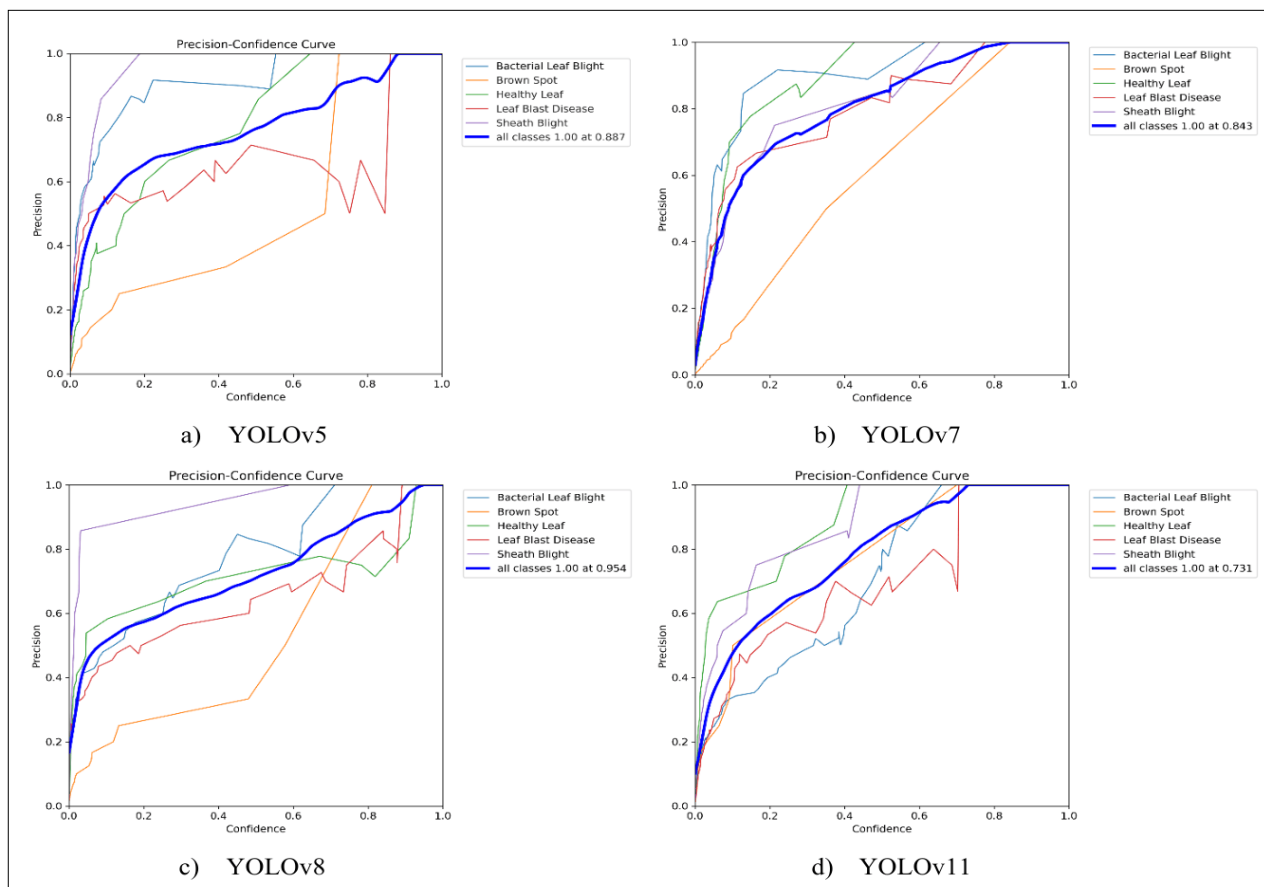


Fig. 9. Precision performance of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models.

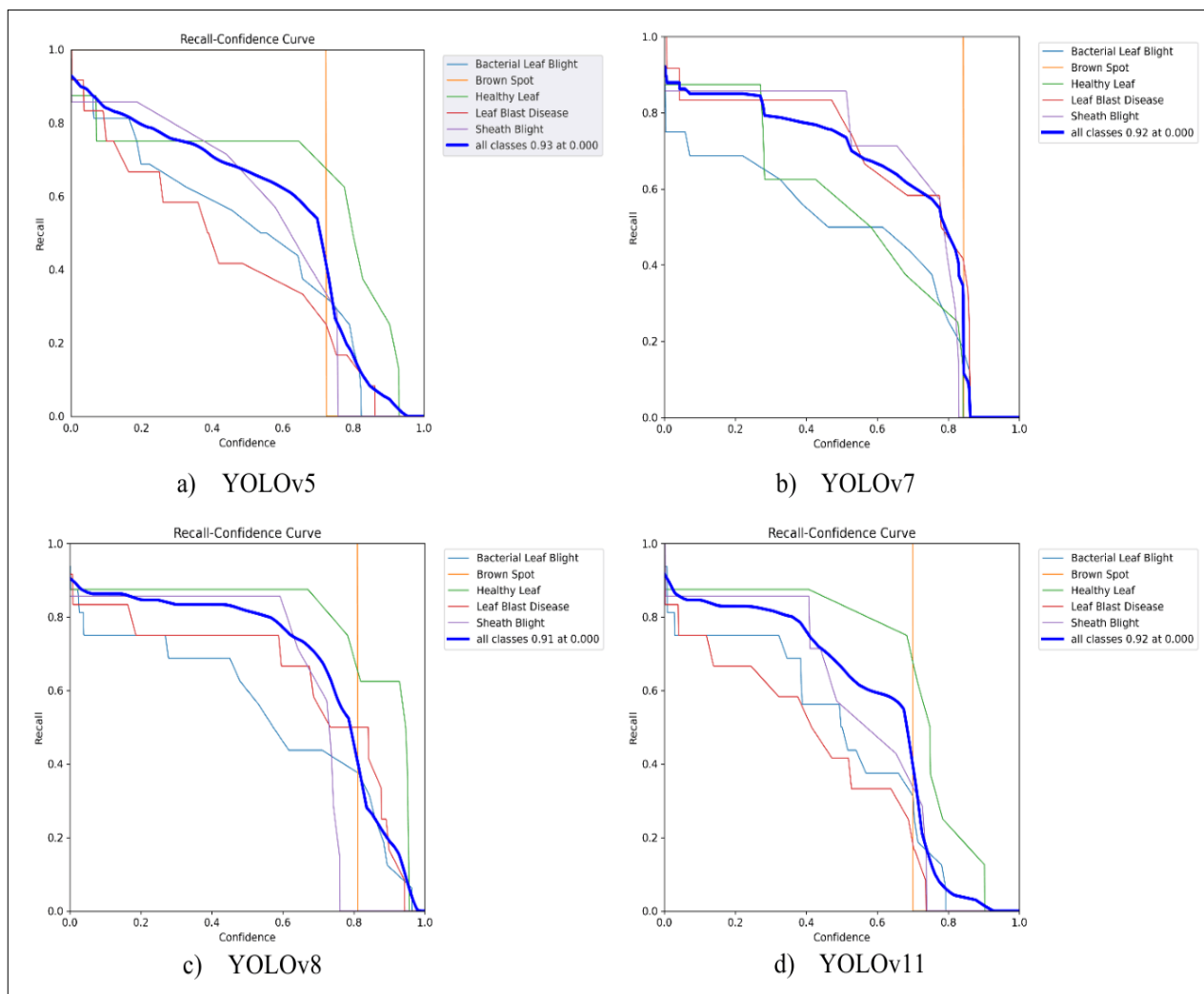


Fig. 10. Recall performance of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models.

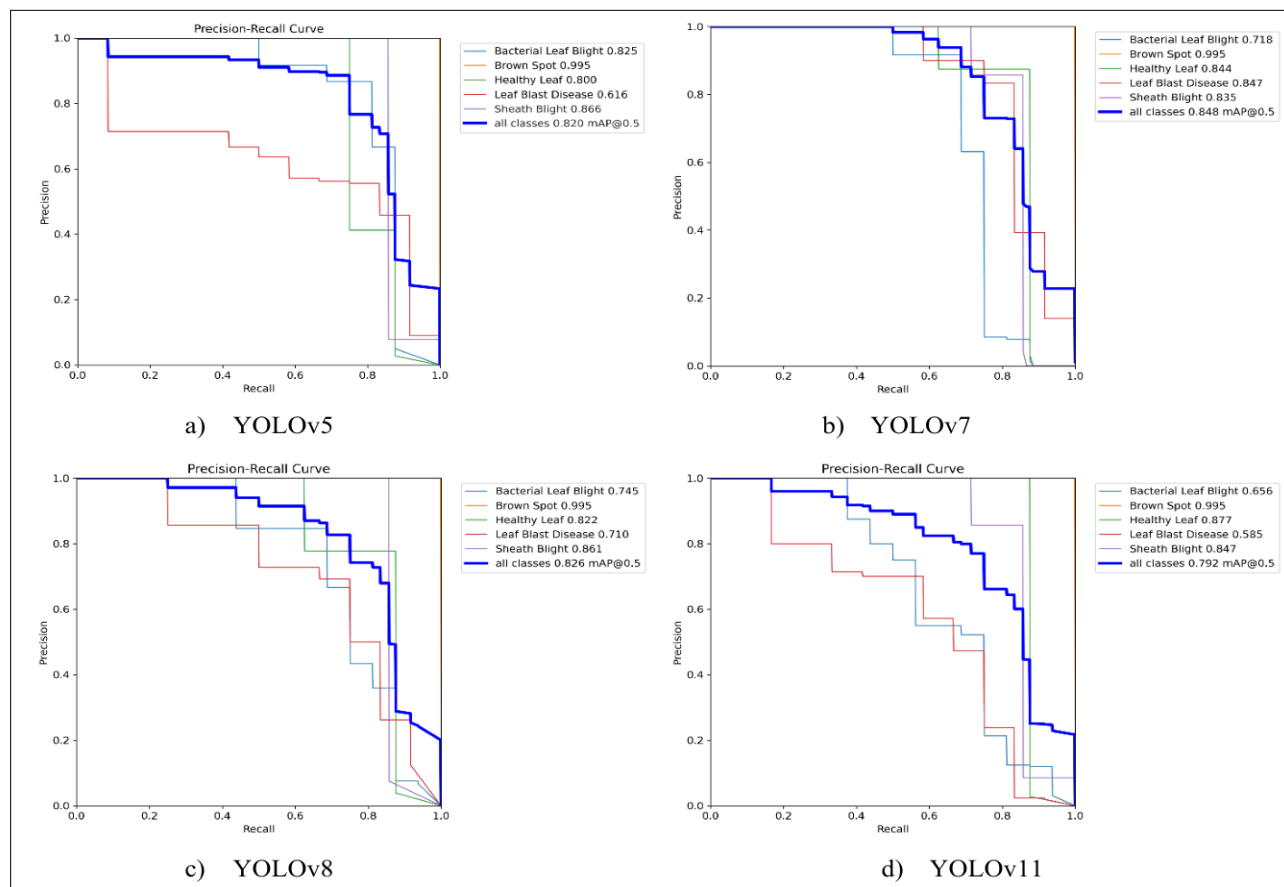


Fig. 11. Precision-recall comparison of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models.

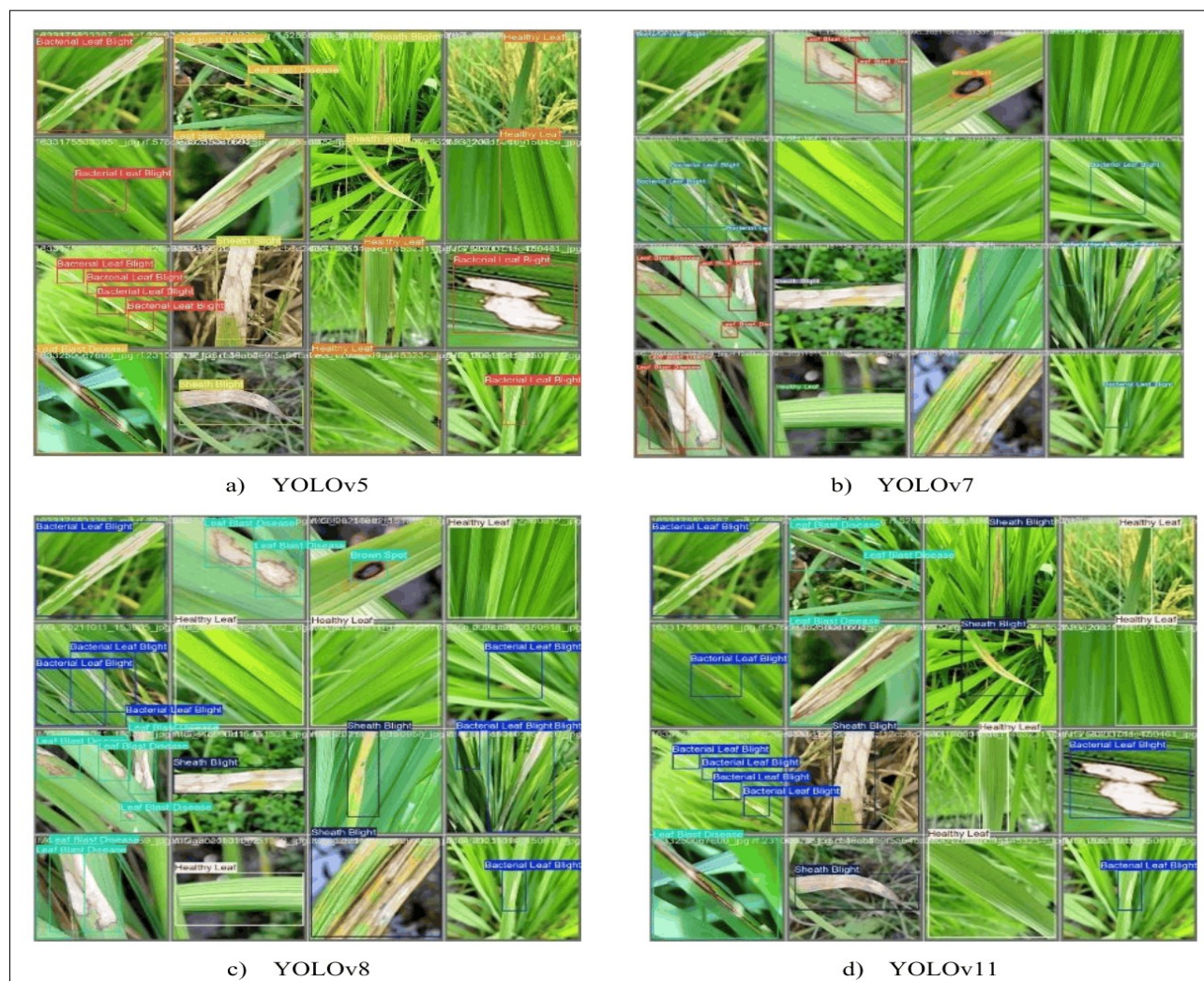


Fig. 12. Subjective performance evaluation of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models training.

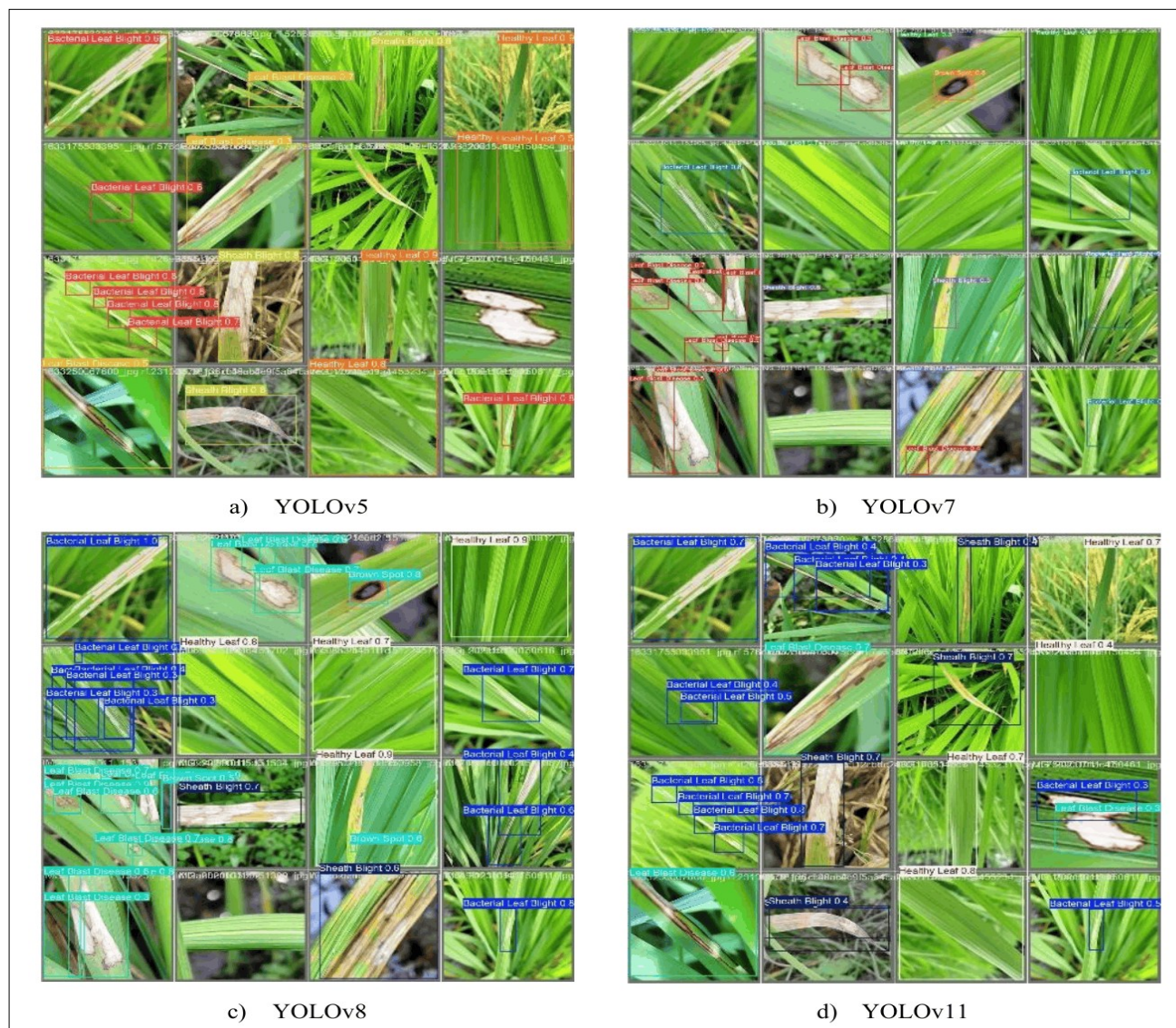


Fig. 13. Subjective performance evaluation of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models testing.

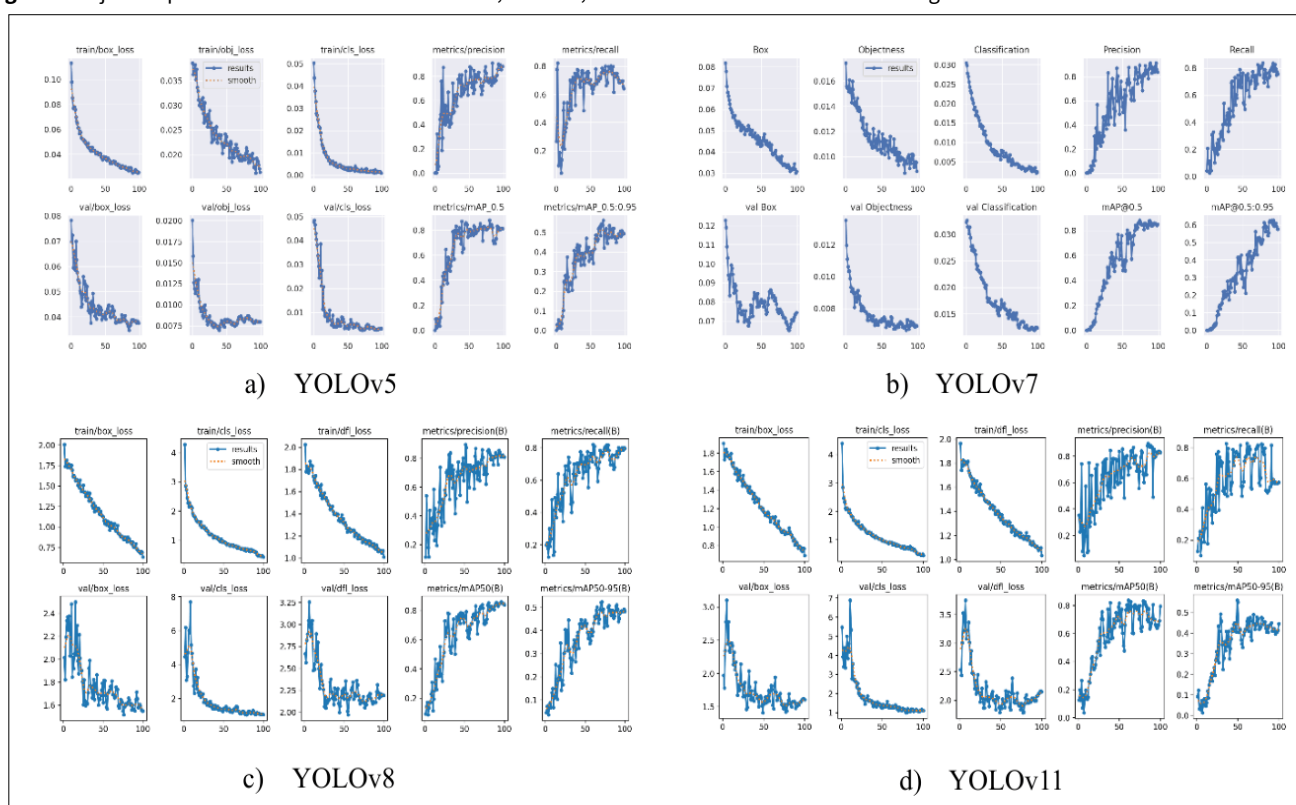


Fig. 14. Training and validation loss curves of YOLOv5, YOLOv7, YOLOv8 and YOLOv11 models.

YOLOv8 and YOLOv11 are shown in Fig. 14. In all the models, it is observed that the box loss, objectless loss and classification loss are on a steady downwards trend, signifying that the networks were able to learn the localisation and classification of rice leaf disease features over the epochs. Model YOLOv7 has shown the most steady and fast convergence amongst the models and it has lesser total loss values and gentler curves, which indicate efficient optimisation and higher generalisation. In contrast, YOLOv8 and YOLOv11 also converge effectively but exhibit more fluctuation in classification loss, suggesting sensitivity to intra-class variability. Although it has a sharp initial decrease in loss, YOLOv5 has relatively higher residual losses than the other models, a fact that is consistent with its lower F1-score and mAP achievements.

Model comparison

The performance of each YOLO model is represented in Table 2. The precision of all models was perfect (1.00) at different confidence thresholds, meaning that when a disease was detected, it was correctly classified. Nonetheless, differences in performance variables like F1 score, recall and mAP have a considerable impact on model effectiveness. Model YOLOv7 has the highest F1 (0.77), mAP (0.85) and accuracy (0.723) values, indicating that it has the best balance between precision and recall. As a result, it provides the most consistent results when dealing with varied situations of the ailment. Model YOLOv5 was the highest recall achiever (0.93) with the lowest F1 score (0.68). It demonstrated a stronger bias for misclassification or overprediction of disease due to its high sensitivity. Although the performances of YOLOv8 were competitive, its recall and mAP scores were slightly lower compared to those of YOLOv7. Therefore, YOLOv7 was more accurate than YOLOv8 in general detection in a rice leaf disease classification problem. Model YOLOv11 had similar F1 and recall values but obtained low mAP compared to YOLOv5, YOLOv7 and YOLOv8. This work comes with several improvements as compared to the studies that have been summarised in Table 1. To start with, even though much of the previous research work is based on small or curated datasets (15, 20), the current research presents a real-world field dataset of 1500 images taken under natural farming conditions, which are more representative of the difficulties of the agricultural environment, such as variable lighting, background clutter and occlusion. Second, most of the previous research only used one version of YOLO (20, 23, 27), but this work presents the first thorough comparison of four different YOLO variants. Like YOLOv5, YOLOv7, YOLOv8 and YOLOv11 trained with the same condition for the purpose of fair benchmarking. Third, the current study does not adopt a multi-metric assessment framework but utilises accuracy in classification; like in previous studies, it uses several metrics, including precision, recall, F1-score, accuracy, mAP and qualitative analysis of visualisation to provide the reader with a better and wider analysis of the performance. Lastly, the findings of the research question may be used to assert that YOLOv7 is the best and most stable solution, which can be applied to the field of performance

related to model parameters, not only to the academic knowledge about the selected instruments but also to the practical conventions of the application of AI devices in the sphere of precision agriculture.

Discussion

The comparative analyses of the four YOLO architectures offer some insights into the appropriateness of the architectures to utilise in real-field scenarios in detecting rice leaf disease. Out of the experimented models, YOLOv7 showed better overall results, which shows that its architecture advances lead to more efficient feature extraction and object localisation in complicated agricultural images. The relatively high recall of YOLOv7 indicates that the model is especially useful when localising diseased leaves, since this is an important aspect in the field of agriculture, where a missed case of disease might result in the loss of a crop. This feature demonstrates the possibility of the model to diagnose and prevent the diseases of crops at the earliest. The differences between the variants of YOLO that could be observed can be explained by the different network designs, feature fusion mechanisms and training optimisation. YOLOv7 incorporates improved training techniques and efficiently designed backbone networks which most probably enhance its accuracy to detect subtle visual patterns of various rice leaf diseases. Contrarily, previous architectures like YOLOv5 might be sensitive to subtle variations in the leaf texture or lesion patterns, particularly with a comparatively small dataset. Although the outcomes achieved are promising, the volume of data and the scope of geographical location is notable areas of weakness. The data used to create figures in this paper was obtained in a particular agricultural setting in Bangladesh and this factor may not be able to extrapolate existing results to additional environmental factors, rice types or disease symptoms. Diets that occur throughout different regions, such as climate differences and soil differences could also affect the appearance of diseases and it is important to train models using more varied and geographically representative data. The results also point to the feasibility of the practical implementation of lightweight object detection models in the field of agricultural monitoring. Since YOLOv7 was very good at such detection, there is an opportunity for having disease-detecting frameworks run on mobile phones or edge computing systems that allow farmers to do the diagnosis themselves without necessarily going to the laboratory or a professional to perform the diagnostic. These AI-based diagnostic aids would help tremendously to improve precision agriculture in both rural and limited-resource farming systems.

Conclusion

The study involved a comparative analysis of 4 you only look once (YOLO) based deep learning (DL) systems to detect and classify rice leaf diseases based on a real-field study that was held in Bangladesh. The dataset was made up of about 1500 annotated images which were subject to five classes: healthy leaves and 4 major rice diseases; namely bacterial leaf blight, brown spot, leaf

Table 2. Model performance comparison

Model	Precision (confidence)	Recall	F1	mAP at 0.5	Accuracy
YOLOv11	1.00 (0.731)	0.92	0.76	0.80	0.62
YOLOv8	1.00 (0.954)	0.91	0.74	0.83	0.61
YOLOv7	1.00 (0.843)	0.92	0.77	0.85	0.723
YOLOv5	1.00 (0.887)	0.93	0.68	0.82	0.66

blast and sheath blight. To guarantee a complete performance comparison, the models were assessed by conventional evaluation tools like precision, recall, F1-score and mean average precision (mAP). The experimental evidence showed that YOLOv7 was better than the other tested models, as it has offered a balanced functionality in relation to detection precision and recall. This suggests that YOLOv7 can efficiently locate disease symptoms in rice leaves and reduce false alarms, a feature that is essential in the detection of early signs of disease in agriculture (especially early disease detection). The results indicate the possibility of the use of advanced object detection models to aid automated crop disease monitoring systems. The efficiency and speed of YOLO-based architectures also demonstrate good potential of real-world deployment of the proposed approach. These systems may be added to mobile or embedded systems, such as edge or internet of things (IoT) devices, so that farmers can diagnose diseases swiftly and conveniently in the field. Such an ability might greatly contribute to precision agriculture activities by enabling quick response and better crop control, especially in the small-scale and farm-scarce rural agricultural settings. However, this study has some limitations. The amount of collected data and its spatial concentration could limit the ability to generalise the trained models to other agriculture types and rice varieties. Further research into work is thus advised to include adding more images taken in different areas and several other rice diseases to make the model robust. Moreover, inclusion of attention mechanisms or lightweight neural networks, including MobileNet-based YOLO versions would enhance computational efficiency whilst retaining precision in detection. Such gains would help improve artificial intelligence (AI) based agricultural disease monitoring systems to be more scalable and reliable.

Acknowledgements

The authors are thankful for the funding by the 2024 Multimedia University (MMU) and King Mongkut's University of Technology North Bangkok (KMUTNB). Matching grant Number (MMUI/240103).

Authors' contributions

AR and EH carried out the rice leaf disease and machine learning studies. AR and CGC participated in the design of the study. AR and EH performed the simulation and code debugging. AR completed the result analysis and drafted the manuscript. LIE conceived the study and participated in its coordination. CGC and LIE are involved in the manuscript revision. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process:

During the preparation of this work, the authors used QuillBot to perform grammar checking and paraphrasing. After using this

tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. U.S. Department of Agriculture, Foreign Agricultural Service. Production. Global Production [Internet]. FAS Data & Analysis; 2025.
2. USDA Foreign Agricultural Service. Bangladesh: Grain and Feed Update [Internet]. 2025.
3. Rice Knowledge Bank. Diseases – IRRI Rice Knowledge Bank [Internet]. 2022.
4. Rashid MM, Nihad SA, Khan MA, Haque A, Ara A, Ferdous T, et al. Pathotype profiling, distribution and virulence analysis of *Xanthomonas oryzae* pv. *oryzae* causing bacterial blight disease of rice in Bangladesh. *J Phytopathol*. 2021;169(7–8):438–46. <https://doi.org/10.1111/jph.13000>
5. Chu J, Li Y, Feng H, Weng X, Ruan Y. Research on multi-scale pest detection and identification method in granary based on improved YOLOv5. *Agriculture*. 2023;13(2):364. <https://doi.org/10.3390/agriculture13020364>
6. Wang A, Peng T, Cao H, Xu Y, Wei X, Cui B. TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field. *Front Plant Sci*. 2022;13:1091655. <https://doi.org/10.3389/fpls.2022.1091655>
7. Simkhada K, Thapa R. Rice blast, a major threat to the rice production and its various management techniques. *Turk J Agric Food Sci Technol*. 2022;10(2):147–57. <https://doi.org/10.24925/turjaf.v10i2.147-157.4548>
8. Chen J, Zhang D, Zeb A, Nanekharan YA. Identification of rice plant diseases using lightweight attention networks. *Expert Syst Appl*. 2021;169:114514.
9. Singh P, Mazumdar P, Harikrishna JA, Babu S. Sheath blight of rice: a review and identification of priorities for future research. *Planta*. 2019;250(5):1387–407. <https://doi.org/10.1007/s00425-019-03246-8>
10. Nayar P, Chhibber S, Dubey AK. An efficient algorithm for plant disease detection using deep convolutional networks. In: *Proc 14th IEEE Int Conf Comput Intell Commun Netw*; 2022; Al-Khobar, Saudi Arabia. p. 156–160. <https://doi.org/10.1109/CICN56167.2022.10008235>
11. Chen J, Zhang D, Nanekharan YA, Li D. Detection of rice plant diseases based on deep transfer learning. *J Sci Food Agric*. 2020;100(7):3246–56. <https://doi.org/10.1002/jsfa.10365>
12. Sethy PK, Barpanda NK, Rath AK, Behera SK. Deep feature based rice leaf disease identification using support vector machine. *Comput Electron Agric*. 2020;175:105527. <https://doi.org/10.1016/j.compag.2020.105527>
13. Bashir K, Rehman M, Bari M. Detection and classification of rice diseases: An automated approach using textural features. *Mehran Univ Res J Eng Technol*. 2019;38(1):239–50. <https://doi.org/10.22581/muet1982.1901.20>
14. Boudaa B, Abada K, Aichouche WA, Belakermi AN. Advancing plant diseases detection with pre-trained YOLO models. In: *Proc 6th Int Conf Pattern Anal Intell Syst*; 2024; El Qued, Algeria. p. 1–6. <https://doi.org/10.1109/PAIS62114.2024.10541267>
15. Krishnamoorthy N, Narasimha Prasad LV, Pavan Kumar CS, Subedi B, Abaha HB, Sathishkumar VE. Rice leaf diseases prediction using deep neural networks with transfer learning. *Environ Res*. 2021;198:111275. <https://doi.org/10.1016/j.envres.2021.111275>
16. Ghosal S, Sarkar K. Rice leaf diseases classification using CNN with transfer learning. In: *Proc IEEE Calcutta Conf*; 2020; Kolkata, India. p. 230–36. <https://doi.org/10.1109/CALCON49167.2020.9106423>
17. Tan L, Lu J, Jiang H. Tomato leaf diseases classification based on

- leaf images: A comparison between classical machine learning and deep learning methods. *AgriEngineering*. 2021;3(3):542–58. <https://doi.org/10.3390/agriengineering3030035>
18. Shahidur Harun Rummy SM, Arefin Hossain MI, Jahan F, Tanvin T. An IoT based system with edge intelligence for rice leaf disease detection using machine learning. In: *Proc IEEE Int IOT Electron Mechatron Conf*; 2021; Toronto, Canada. p. 1–6. <https://doi.org/10.1109/IEMTRONICS52119.2021.9422499>
 19. Yao Q, Guan Z, Zhou Y, Tang J, Hu Y, Yang B. Application of support vector machine for detecting rice diseases using shape and color texture features. In: *Proc Int Conf Eng Comput*; 2009; Hong Kong, China. p. 79–83. <https://doi.org/10.1109/ICEC.2009.73>
 20. Pallathadka H, Ravipati P, Sajja GS, Phasinam K, Kassanuk T, Sanchez DT, et al. Application of machine learning techniques in rice leaf disease detection. *Mater Today Proc*. 2022;51:2277–80. <https://doi.org/10.1016/j.matpr.2021.11.398>
 21. Agbulos MK, Sarmiento Y, Villaverde J. Identification of leaf blast and brown spot diseases on rice leaf with YOLO algorithm. In: *Proc 7th Int Conf Control Sci Syst Eng*; 2021; Qingdao, China. p. 307–12. <https://doi.org/10.1109/ICCSSE52761.2021.9545153>
 22. Mia J, Bijoy HI, Uddin S, Raza DM. Real-time herb leaves localization and classification using YOLO. In: *Proc Int Conf Comput Commun Netw Technol*; 2021; Kharagpur, India. p. 1–7. <https://doi.org/10.1109/ICCCNT51525.2021.9579718>
 23. leamsaard J, Charoensook SN, Yammen S. Deep learning-based face mask detection using YOLOv5. In: *Proc Int Electr Eng Congr*; 2021; Pattaya, Thailand. p. 428–31. <https://doi.org/10.1109/IEECON51072.2021.9440346>
 24. Kundu N, Rani G, Dhaka VS. Seeds classification and quality testing using deep learning and YOLOv5. In: *Proc Int Conf Data Sci Mach Learn Artif Intell*; 2022. p. 153–60. <https://doi.org/10.1145/3484824.3484913>
 25. Haque ME, Rahman A, Junaid I, Hoque SU, Paul M. Rice leaf disease classification and detection using YOLOv5. *arXiv*. 2022;arXiv:2209.01579.
 26. Haque E, Paul M, Rahman A, Tohidi F, Islam J. Rice leaf disease detection and classification using shallow trained YOLOv7 active deep learning approach. In: *Proc Int Conf Digit Image Comput Tech Appl*; 2023; Port Macquarie, Australia. p. 516–22. <https://doi.org/10.1109/DICTA60407.2023.00077>
 27. Jubayer F, Soeb JA, Mojumder AN, Paul MK, Barua P, Kayshar S, et al. Detection of mold on the food surface using YOLOv5. *Curr Res Food Sci*. 2021;4:724–28. <https://doi.org/10.1016/j.crfs.2021.10.003>
 28. Zhou F, Zhao H, Nie Z. Safety helmet detection based on YOLOv5. In: *Proc IEEE Int Conf Power Electron Comput Appl*; 2021; Shenyang, China. p. 6–11. <https://doi.org/10.1109/ICPECA51329.2021.9362711>
 29. Chen Z, Wu R, Lin Y, Li C, Chen S, Yuan Z, et al. Plant disease recognition model based on improved YOLOv5. *Agronomy*. 2022;12(2):365. <https://doi.org/10.3390/agronomy12020365>
 30. Mohandas A, Anjali MS, Rahul Varma U. Real-time detection and identification of plant leaf diseases using YOLOv4-tiny. In: *Proc Int Conf Comput Commun Netw Technol*; 2021. p. 14628–38. <https://doi.org/10.1109/ICCCNT51525.2021.9579783>
 31. Yao X, Yang F, Yao J. YOLO-Wheat: A wheat disease detection algorithm improved by YOLOv8s. *IEEE Access*. 2024;12:133877–888. <https://doi.org/10.1109/ACCESS.2024.3460394>
 32. Jia L, Wang T, Chen Y, Zang Y, Li X, Shi H. MobileNet-CA-YOLO: An improved YOLOv7 based on MobileNetV3 and attention mechanism for rice pests and diseases detection. *Agriculture*. 2023;13(7):1285. <https://doi.org/10.3390/agriculture13071285>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.