



OPEN Classification of cotton leaf disease using YOLOv8 based k-fold cross validation deep learning method for precision agriculture

Kamaldeep Joshi¹, Yashasvi Yadav¹, Sahil Hooda¹, Rainu Nandal¹, Baljinder Singh², Kashmir Singh³✉, Narendra Tuteja⁴, Ritu Gill⁵✉ & Sarvajeet Singh Gill⁵✉

Cotton production is a crucial agricultural industry, a raw material source for the textiles sector and a major source of livelihood for more than 30 million farmers globally. The yield and quality of cotton (*Gossypium*) are influenced by different types of stress and diseases. Deep Learning as a solution for disease prevention, detection, and management can increase the yield, reduce the cost and improve the quality of crop. This study presents a robust method using 10-fold cross-validation with the YOLOv8 DL model for precise cotton leaf disease recognition. The k-fold cross-validation mitigates overfitting by training the model on diverse data subsets, which leads to enhanced generalizability while ensuring reliable performance. The proposed method achieved 99.60% and 100% as Top_1 and Top_5 accuracy, respectively. The method also achieved a recall of 99.53%, a precision of 99.53%, and an F1 score of 99.60%. During 10 trials, the method consistently performed with an average. Top_1 and Top_5 accuracy of 98.41% and 100% respectively, recall 98.53%, precision 98.39% and F1 score 98.42%. This study is among the first to apply YOLOv8 classification with 10-fold cross-validation for multi-class cotton leaf disease identification using field-captured images.

Keywords Cotton leaf disease recognition, Deep learning, Disease classification, K-fold cross-validation, Precision agriculture, YOLOv8 deep learning model

Precision agriculture has become increasingly significant in addressing challenges related to global food security, environmental sustainability, and economic efficiency^{1–3}. Precision agriculture offers timely and accurate diagnosis of plant diseases, which can significantly reduce crop losses and enhance yield quality^{4,5}. Cotton (*Gossypium*) is an essential cash crop and a cornerstone of the textile industry, providing raw materials for clothing and fabric production^{6,7}. It is particularly important in countries like Pakistan, Bangladesh, and India, where it serves as a major economic driver^{8,9}. In Pakistan, cotton contributes nearly 10% to the GDP and accounts for 55% of foreign exchange earnings, with approximately 1.5 million people engaged in its value chain⁸. Similarly, India cultivates 24% of the world's cotton-growing land, generating substantial revenue from crops⁹. Unlike synthetic fibres such as polyester and nylon, which are less environmentally friendly, cotton is biodegradable and can improve soil health when managed sustainably⁹. However, the crop is highly susceptible to various biotic and abiotic stresses, including bacterial, viral, and pest-induced diseases, which can cause severe economic losses⁷. The process, speed and cost of these stress detection and management is a major influence on crop yield and quality^{9,10}. Recent advancements in artificial intelligence (AI) and deep learning (DL) have transformed the agricultural sector, leading to the development of automated systems for recognizing plant diseases^{10–14}. Among these advancements, the You Only Look Once (YOLO) architecture has become particularly well-known for its speed and accuracy in object detection and classification tasks^{15,16}. The latest YOLOv8 model features improved capabilities for precise and efficient classification, making it an excellent choice for diagnosing cotton leaf diseases across various environmental conditions¹⁵. Automated systems that utilize DL enable real-time monitoring and data analytics, allowing farmers and researchers to identify issues early and take corrective

¹Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, Haryana, India. ²Department of Biochemistry, School of Basic Science, Central University of Punjab, Bathinda 151401, India. ³Department of Biotechnology, Panjab University, Chandigarh, India. ⁴Plant Molecular Biology Group, International Centre for Genetic Engineering & Biotechnology (ICGEB), New Delhi 110 067, India. ⁵Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India. ✉email: kashmirbio@pu.ac.in; ritu_gill@mdurohtak.ac.in; ssgill14@mdurohtak.ac.in

actions^{17,18} These systems analyse spectral signatures to evaluate and classify cotton plants, offering insights into crop diseases, pests, and environmental stressors. Ultimately, this improves crop management and optimizes production⁹.

Many different DL models are prevalent for real-time disease detection in cotton plants, which are mentioned in Table 1. CDDLite-YOLO model is one such model achieving an average precision of 90.6% with

References	Objectives	Dataset used	Results	Limitations
Elaraby et al., 2022b ²⁰	DL model for stress detection in five different crops, i.e., cucumber, corn, wheat, grape, cotton	54k images of fourteen different crops from Plant Village dataset	Accuracy 98.83%, Sensitivity (sens.) 98.78%, F Score 98.47%, Precision (precision) 98.67%, and Specificity (spec.) 98.53%	Real-time implementation testing required
Pan et al., 2024 ²¹	CDDLite-YOLO model for cotton plant stress detection + enhancing acc. with minimum YOLO parameters	1530 natural field images of cotton plant (38% verticillium wilt, 34% anthracnose, and 28% fusarium wilt)	(Mean Average Precision) mAP achieved is 90.60% with 3.6G FLOPS and 1.8 M parameters + identification speed of 222.22 FPS	Only 78.10% mAP for verticillium detection
Ahmed, 2021 ²²	Transfer learning based Custom CNN i.e., DCPLD-CNN for cotton leaf disease detection.	The dataset used in the Cotton Plant and Leaf Disease recognition study was collected from a specified source [38]	Accuracy 98.77% + validation acc. of 88.99% and 98.77% for 100 and 500 iterations, respectively	More robust testing specific to cotton plant dataset needed
Gao et al., 2024 ²³	DL model for cotton plant stress detection	Datasets for the study were collected from diversified sources through manual collection and internet crawling techniques	94% accuracy, 95% mAP, and speed of 49.7 FPS	Further study is needed to maintain performance on larger datasets and reduce computational resource consumption
Bharathi et al., 2024 ²⁴	Random tree based- adaptive fire-hawk DL model i.e., DQRR-AFH	Hybrid database by combining 1710 natural field and internet images of cotton plant	98.88% accuracy, 99.21% F1 score, 97% precision + Performance comparison with WL-CNN, ECPRC, and DT models	Only two class classifications are performed
Li et al., 2024 ²⁵	CFNet-VoV-GCSP-LSKNet-YOLOv8s model for cotton stress detection	6 public datasets from Kaggle	89.9% precision, 90.70% recall, and 93.7% mAP(0.5)	The study does not address the limitations explicitly, focusing more on the proposed method's enhancements and performance metrics.
Nazeer et al., 2024 ²⁶	Develop a dataset of cotton leaf images to support automated disease detection systems + DL model for detection of Cotton Leaf Curl Disease (CLCuD)	Hybrid dataset of Kaggle images and natural self-collected 1349 images of cotton leaf	99% accuracy achieved	Model restricted to Leaf curl disease in cotton plant
Latif et al., 2021 ⁸	Develop an automated technique for detecting cotton leaf diseases using DL	The study utilized 1000 self-collected datasets of cotton diseases labelled and augmented by an expert for training and testing purposes	Achieved an accuracy of 98.8% using Cubic SVM	The model applied to four classes only, which are Areolate Mildew, Myrothecium leaf spot, and Soreshine
Kolachi et al., 2023 ²⁷	Identification of blight and curl disease in cotton plant using a custom YOLO DL model	Natural dataset from farmer's fields of Sindh, Pakistan. It consists of healthy leaves, bacterial blight, and curl virus images. 1000 images were sourced from Kaggle, GitHub, and Google to enhance the dataset's diversity and size. After augmentation, a total of 5046 images were formed	The YOLOv5 model achieved 92% accuracy in disease classification	Only two class classifications are performed: bacterial blight and curl virus
Zhu et al., 2022 ²⁸	Develop a cotton disease identification method based on pruning on VGG16, ResNet164, and DenseNet40 to address deployability issues on resource-limited smart devices	PlantVillage consists of 14 types of plants with 54,306 images of healthy and diseased leaves	The compressed models (size 2.2 MB) achieved high accuracies, with DenseNet40-80-T achieving 97.23%	Lack of detailed discussion on the potential challenges or drawbacks of the proposed pruning algorithm and compression strategies used for cotton disease identification based on DCNN models
Thivya et al., 2024 ⁹	Develop a novel DL pipeline, CoDet, for cotton plant detection and disease identification	30k images collected from internet (25k for training and rest 5k for validation testing)	CoDet outperformed other models in comparative study using different matrices	More robust testing needed considering Indian environment and physiochemical traits of cotton plant
Rai and Pahuja, 2023 ²⁹	Deep-CNN for cotton stress detection	Hybrid dataset of 2293 natural images and Kaggle images	97.98% accuracy achieved	Real time performance testing needed
Rai and Pahuja, 2024 ³⁰	Using DL methods for cotton stress identification	Two different datasets from Kaggle of 2310 and 1711 images respectively	99.48% accuracy and 99% sens. achieved	More robust testing needed
Shahid et al., 2024 ³¹	GoogleNet, VGG19, AlexNet, and InceptionV3 for identification of cotton plant stress	Natural dataset from Balochistan, Pakistan covering all the 3 phases i.e., sowing, germination and maturity	Accuracy achieved by GoogleNet, AlexNet, and InceptionV3 is 93.40%, 93.40%, and 91.80% respectively	The accuracy is low as compared to the other models
Kukadiya et al., 2024 ¹⁰	Pre-trained VGG16 and InceptionV3 used to detect early cotton leaf diseases	1786 images of 4 different types i.e., blight, curl, wilt and healthy from PlantVillage dataset	Training and Testing accuracy of 98% and 95% achieved respectively	It focused only on four cotton diseases
Islam et al., 2023 ¹⁹	Hybrid DL models by combining Transfer learning with Xception, Inception V3, VGG 16 and 19	2310 images from Kaggle dataset	VGG-16 achieved an accuracy of 90.22%, while VGG-19, Inception-V3, and Xception achieved higher accuracies of 96.74%, 97.83%, and 98.70%, respectively	Only binary classification achieved, more robust testing needed

Table 1. The following table explores the study in terms of author and year of publication (reference), objectives of the study, dataset used for the study, results of the study and limitations of the study.

easy deployment on resource-constrained devices. These advancements ensure timely disease detection and intervention, which are crucial for maintaining cotton yield and quality¹⁰. Additionally, techniques such as model pruning minimize computational overhead, allowing deployment on mobile devices without sacrificing accuracy. These advancements enable farmers to proactively tackle crop issues, leading to improved yield optimization⁹.

This study presents a systematic workflow for identifying and classifying cotton leaf diseases using the YOLOv8m classification model. The dataset used in this study is a high-resolution “SAR-CLD 2024” image dataset. This dataset consists of seven categories of leaf images, *i.e.*, healthy, herbicide-infected, leaf hopper jassids, bacterial blight, red leaf, curl virus, and variegated leaves. Preprocessing is integrated before k-fold cross-validation, ensuring higher reliability and robustness of the model in diverse conditions. The following objectives are identified for this study:

1. To identify the area of research that includes the AI-based diagnosis of cotton leaf diseases.
2. To utilize the YOLOv8 deep learning architecture to accurately classify multiple cotton leaf diseases using real-field images.
3. To implement a k-fold cross-validation approach to reduce overfitting, improve robustness, and ensure the model performs consistently across diverse subsets of data.
4. To achieve high model performance, ensuring reliable and balanced disease classification, which minimizes false predictions.

By utilizing advanced DL techniques, the proposed system has significant potential to improve crop management practices and alleviate the negative impacts of cotton diseases on crop performance¹⁹. Furthermore, this study thoroughly assesses the performance of the model, establishing a foundation for future innovations in automated plant disease detection systems.

Recent developments in DL-assisted disease detection in plants

Recent advancements in the detection of cotton leaf disease and machine vision classification, as shown in Table 1, have been significant. Search Query (“Cotton” AND “Deep Learning”) has been defined for extracting relevant studies from Frontiers, Web of Science, Science Direct, IEEEExplore, and Springer Link databases. The initial findings showed that there were limited publications specifically focused on disease detection in cotton leaves. Table 1 summarizes the studies, highlighting the authors, publication years, study objectives, the dataset used, results, and identified limitations.

The above study concluded that most approaches identified and classified a maximum of four classes. While these models achieved good accuracy, their effectiveness was limited due to the few classes in the dataset. Additionally, most studies relied on a single DL model. To the best of our knowledge, no prior study has applied k-fold cross-validation specifically with YOLO-based architectures, particularly YOLOv8, for multi-class cotton leaf disease classification using field images. Our approach overcomes these limitations and produces a robust, high-accuracy model to mitigate them.

Materials and methods

The proposed work follows the workflow shown in Fig. 1. It starts with collecting data from the “SAR-CLD-2024” (<https://data.mendeley.com/datasets/b3jy2p6k8w/2>) dataset³² which contains images categorized into seven classes, of diseases and healthy leaves. During the pre-processing stage, the dataset is resized and organized into a standardized format suitable for classification. The workflow uses k-fold cross-validation, which divides the dataset into multiple folds to ensure robust training and evaluation. The YOLOv8m architecture is employed for neural network training to classify the images. Finally, the process includes a validation phase, where the predictions are assessed for performance.

Dataset and preprocessing

The dataset was sourced from the SAR-CLD-2024 dataset, which contains high-quality images of cotton disease. Dataset of 2137 images from the NCRI (National Cotton Research Institute), Gazipur. The images are taken by a smartphone (Redmi Note11s). This robust dataset covered 7 different classes, including both biotic and abiotic stresses.

The leaves from all 7 classes are illustrated in Fig. 2, and the names of the cotton diseases and their corresponding images are shown in Table 2.

To apply the YOLO classification model to the obtained dataset, the data must be organized into three folders: “train”, “val”, and “test”. Each folder contains seven subfolders, each named after one of the seven classes, with the corresponding images for that class. Figure 3 illustrates the format used by YOLO to classify the dataset.

Several preprocessing steps are applied to ensure consistency in the model’s learning efficiency. All images were resized to 640 × 640 pixels, matching the input size of the YOLOv8 architecture. The dataset was split using Python, with the data randomly divided into training (69%), validation (12%), and testing (19%) sets. A significant number of images were allocated for testing to assess the accuracy of the trained model. Table 3 outlines the distribution of the dataset into training, validation, and testing subsets for each individual class.

K-fold validation

This technique is applied to divide the dataset into ‘k’ parts, named ‘folds’, to carry out a more accurate method of model performance. Each fold provides data for both training and validation. The k-fold cross-validation applied to object classification scenarios ensures robustness to the DL model, making it perfectly generalizable for various data splits. Cross-validation is particularly important in agriculture, as environmental conditions

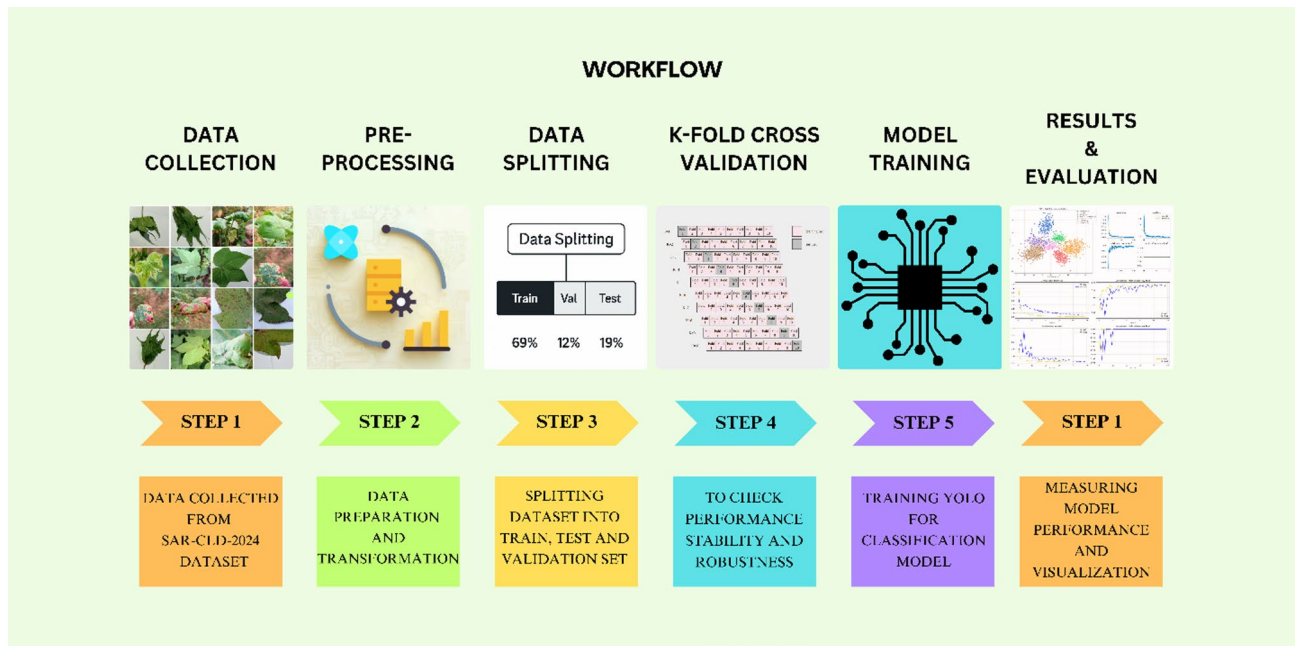


Fig. 1. Schematic workflow of the research work.

may vary, causing the appearance of leaves and disease symptoms to differ from those seen in the training set^{33,34} (Sohail et al., 2023; Samuel et al., 2024). K-fold cross-validation, combined with DL architectures such as CNN and ResNet-152V2, has been shown to improve the predictive capabilities of the model for classifying and diagnosing cotton plant diseases, thus enhancing its effectiveness in real-world applications (Jai Vignesh et al., 2023)³⁵. Training dataset frequently suffers from overfitting, *i.e.*, reduced performance on new, unknown images. K-fold cross-validation addresses this problem by evaluating the performance of the model across different data partitions. It ensures that the model does not simply memorize the training data but instead learns to generalize (Gayatri et al., 2022)³⁶. For further enhancement of the model's robustness, a largely diversified dataset for cross-validation is used³⁴ (Samuel et al., 2024; Kumar et al., 2024)^{34,37}. To further strengthen the model, we employed the k-fold technique, creating ten distinct training, validation, and test folds. Each fold was randomly split to ensure variability in the dataset, with the random splitting and fold formation implemented using Python programming. An example of a k-fold process is shown in Fig. 4. The dataset is split into three categories, and the same process is done for ten different Folds. For each of the ten folds, the dataset was divided into three categories, with each fold containing a unique set of images.

Experimental setup

The output of an image classifier consists of a single class label and a confidence score. Image classification is particularly useful when the goal is to identify the class to which an image belongs without needing to pinpoint the exact location or shape of the objects within it. YOLOv8 models, specifically the yolov8m-cls.pt variant (Fig. 5) is designed for efficient image classification. The model assigns a class label and a confidence score to an entire image. This approach is especially valuable in applications where knowing the class of an image is sufficient, rather than requiring detailed information about the location or shape of objects it contains.

The YOLOv8m-cls model contains 141 layers, 15,781,303 parameters, 15,781,303 gradients, and 41.9 GFLOPs. Out of these, we used 103 layers, 15,771,623 parameters, 0 gradients, and 41.6 GFLOPs. NVIDIA GeForce RTX 3050 Ti Laptop GPU, 4096MiB and Intel i7 12th gen processor were used to perform the desired experiment. Initial hyperparameters {lr=0.01, momentum=0.937, lrf=0.01, wgt_decay=0.0005, warmup_epochs=0.0005, warmup_decay=3.0, warmup_momentum=0.8, warmup_bias_lr=0.1, box=7.5, cls=0.5, dfl=1.5, pose=12.0, kobj=1.0, label_smoothing=1.0, label_smoothing=0.0, and nbs=64} have been used.

Figure 6 illustrates the augmentation strategies of the YOLOv8 model. Default parameters {hsv_h=0.015, hsv_s=0.7, hsv_v=0.4, degrees=0.0, translate=0.1, scale=0.5, shear=0.0, perspective=0.0, flipud=0.0, fliplr=0.5, bgr=0.0, mosaic=1.0, mixup=0.0, copy_paste=0.0, auto_augment: randaugment, erasing=0.4 and crop_fraction=1.0} has been used.

These augmentation techniques address the class imbalance present in the SAR-CLD-2024 dataset. It increases the representation of minority classes and helps the model to learn balanced features, which improves the generalization and reduces class-wise prediction bias. Additionally, the use of 10-fold cross-validation ensured, all classes were fairly represented across training and validation splits.

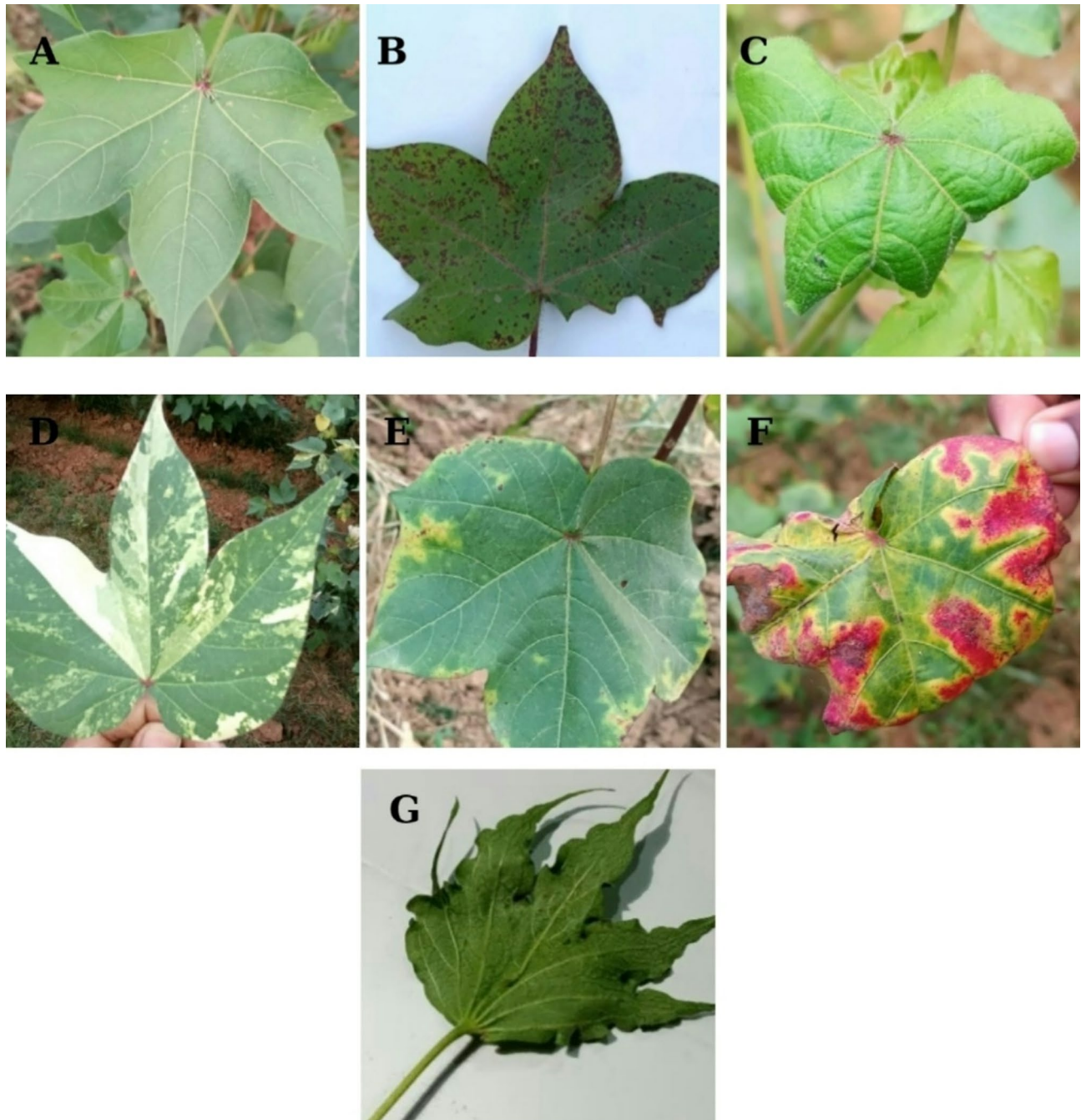


Fig. 2. The individual image of the seven classes. (A) Healthy Leaf, (B) Bacterial Blight, (C) Curl Virus, (D) Leaf Variegation, (E) Jassids by Leaf Hopper, (F) Red Leaf and (G) Herbicide Growth Damage.

Results and validation

The model has been thoroughly tested and evaluated using a wide variety of matrices. The main metrics used are: precision, recall, F1 score, and mean Average Precision (mAP). The fundamental principles of two positives, *i.e.*, True Positive (T.P.) and False Positive (F.P.) and two negatives, False Negative (F.N.) and False Positive (FP), have been used for the calculation of metrics.

- **Accuracy** is evaluated by calculating the percentage of correct predictions as a ratio of total predictions.

$$Accuracy = \frac{T.P. + T.N.}{T.P. + F.P. + T.N. + F.N.} * 100 \quad (1)$$

Sr. No.	Disease name	Number of images
A	Healthy Leaf	257
B	Blight (bacteria)	250
C	Curl (virus)	431
D	Leaf Variegation	116
E	Jassids by Leaf Hopper	225
F	Red Leaf	578
G	Herbicide Growth Damage	280
	Total	2137

Table 2. Number of images in individual class.

```

Original Dataset/
|-- train/
|   |-- Healthy Leaf/
|   |-- Bacterial Blight/
|   |-- Curl Virus/
|   |-- Leaf Variegation/
|   |-- Leaf Hopper Jassids/
|   |-- Leaf Redding/
|   |-- Herbicide Growth Damage/
|-- test/
|   |-- Healthy Leaf/
|   |-- Bacterial Blight/
|   |-- Curl Virus/
|   |-- Leaf Variegation/
|   |-- Leaf Hopper Jassids/
|   |-- Leaf Redding/
|   |-- Herbicide Growth Damage/
|-- val/
|   |-- Healthy Leaf/
|   |-- Bacterial Blight/
|   |-- Curl Virus/
|   |-- Leaf Variegation/
|   |-- Leaf Hopper Jassids/
|   |-- Leaf Redding/
|   |-- Herbicide Growth Damage/

```

Fig. 3. Format of dataset for YOLO classification.

- **Precision** is evaluated by calculating the percentage of correct positive predictions as a ratio of all positive predictions.

$$\text{Precision} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.P.}} * 100 \quad (2)$$

- **Recall** is evaluated by calculating the percentage of true positives as a ratio of all real positives.

$$\text{Recall} = \frac{\text{T.P.}}{\text{T.P.} + \text{F.N.}} * 100 \quad (3)$$

- **F1 Score** is evaluated by calculating the Harmonic Mean of precision and recall.

Disease name	Train	Validation	Test	Total
Bacterial blight	172	31	47	250
Curl virus	297	53	81	431
Healthy leaf	176	31	50	257
Herbicide growth damage	192	34	54	280
Leaf Hopper Jassids	155	28	42	225
Leaf redding	398	72	108	578
Leaf variegation	79	14	23	116
Total	1469	263	405	2137

Table 3. Distribution of images in train, validation, and test.

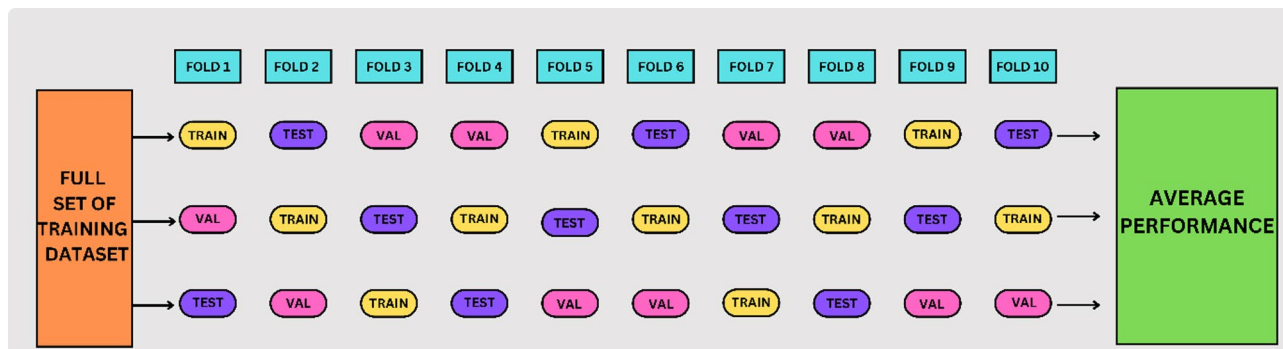


Fig. 4. K-fold splitting of the dataset.

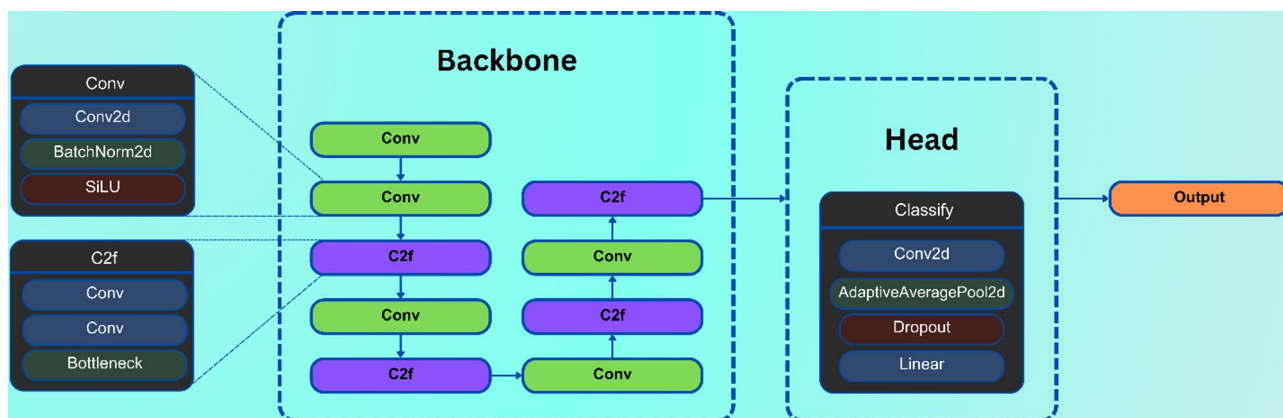


Fig. 5. Shows the detailed architecture of the YOLOv8 classification model.

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{1 * \text{Precision} + \text{Recall}} * 100 \quad (4)$$

- **Mean average precision (mAP):** mAP is the mean of the Average Precision (AP) across all classes, where AP is the area under the precision-recall curve.

$$\text{mAP} = \frac{1}{n} \sum_{k=1}^n \text{AP}(n) \quad (5)$$

mAP is typically evaluated at different IoU thresholds, such as 50% (mAP50) and between 50% and 95% (mAP50-95).

- **mAP50 (B)** This is the mAP calculated specifically for bounding box detection at an IoU threshold of 50%.



Fig. 6. Augmentations of YOLO model shows the different types of augmentations used internally by the YOLOv8m classify model to classify the leaf for example the default settings are {hsv_h=0.015, hsv_s=0.7, hsv_v=0.4, degrees=0.0, translate=0.1, scale=0.5, shear=0.0, perspective=0.0, flipud=0.0, fliplr=0.5, bgr=0.0, mosaic=1.0, mixup=0.0, copy_paste=0.0, auto_augment: randaugment, erasing=0.4, and crop_fraction=1.0}.

Evaluation of YOLOv8

Figure 7 illustrates the plots depicting losses (both training and validation) and Top_1 and Top_5 accuracy for 100 epochs. The losses decreased and stabilized at 0.1 and 1.2 for training and validation, respectively. The value of these losses demonstrates that effective learning with minimum overfitting is achieved. The Top_1 accuracy exhibits a rapid increase from approximately 80% to around 99%, demonstrating the strong ability of the model to predict the correct class on the first attempt. The Top_5 accuracy remains consistently at 1.0, signifying that the model consistently includes the correct label within its Top_5 predictions. Table 4 illustrates trial metrics, and Table 5 illustrates best trials.

The 100% Top-5 accuracy achieved by the model is expected in this case due to the limited number of classes (8) and strong performance of the trained model. Since Top-5 accuracy only checks whether the correct label appears in the top five predictions, such results are common when the model learns well-separated features. However, Top-1 accuracy remains the primary indicator of model effectiveness, as it reflects the model's ability to correctly predict the disease in a single attempt.

A consistently high performance with 98.41% accuracy, 98.39% precision, 98.53% recall, and 98.42% of F1_Score is achieved throughout the 10 trials as illustrated in Table 4. The second trial yielded the best results, achieving the highest accuracy at 99.60%, while the other trials also had strong performance. This suggests that the model is robust (Fig. 10), with minor variations in the results likely due to differences in conditions (Figs. 11, 12, 13, 14).

Despite achieving high accuracy values (Top-1: 99.60%, Top-5: 100%), the proposed model does not suffer from overfitting. This conclusion is supported by multiple observations drawn from model behavior and data characteristics. Firstly, the model was evaluated using 10-fold cross-validation, ensuring that each subset of data is used for both training and validation. The performance remained consistent across all folds, which reflects the robustness and generalizability of the model. Secondly, the confusion matrix generated during validation shows minimal misclassifications, which confirms that the model maintains its classification abilities on unseen data. Also, the SAR-CLD-2024 dataset, without any augmentations, containing real-world unique images, is used to train the model. No synthetic data or repetition was used during training, which guarantees that the model has learnt diverse and realistic field conditions.

Principal Component Analysis (PCA) was performed on the deep feature vectors extracted from the final layer of the YOLOv8 classifier. As shown in Fig. 8, the embeddings from different classes have formed distinct and well-separated clusters in the plot. This evidence confirms that the model has effectively learnt discriminative features and is not merely memorizing the training data.

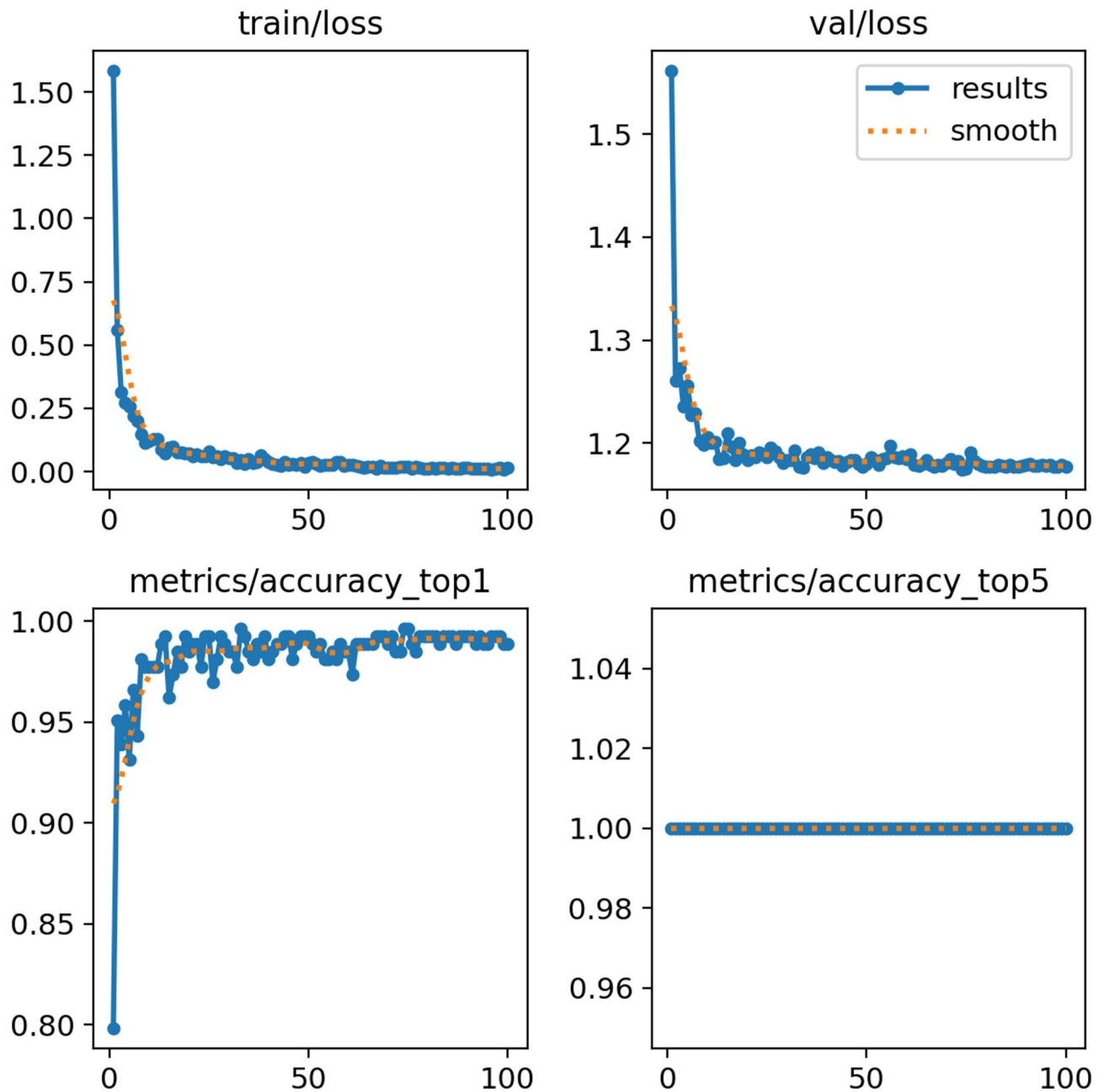


Fig. 7. Best trial validation results show the four graphs, including two metrics and two loss graphs. All four graphs show excellent results; both losses are equivalent to zero. Top_1 and Top_5 accuracy of 99.60% and 100% is achieved respectively.

Comparative evaluation of YOLOv8 with YOLOv11

To further evaluate the effectiveness of the proposed YOLOv8 model, a comparative analysis with YOLOv11, which is a recently released version of the YOLO architecture, is conducted. Both models were trained and validated on the same dataset using identical parameters, including batch size, epochs, and input resolution.

As shown in Fig. 9, YOLOv8 consistently outperformed YOLOv11 in key performance metrics, including train_loss, val_loss, top1 accuracy and top5 accuracy. This suggests that although YOLOv11 is a newer version in the YOLO series, it may not yet be fully optimised for image classification tasks, particularly in the context of fine-grained agricultural disease detection.

Comparison of classification metrics between YOLOv8 and YOLOv11 on cotton leaf disease classes. YOLOv8 demonstrates smoother convergence, supporting its use for the proposed method. Our experiments revealed that YOLOv11 struggled to achieve stable convergence, as shown in Fig. 9, with fluctuating loss curves and lower accuracy. In contrast, YOLOv8 offers a well-balanced architecture and consistent results across multiple datasets, especially in our use case, making it more suitable for deployment in real-world agricultural scenarios. It is also important to note that YOLOv9 and YOLOv10 do not provide support for image classification tasks, which

Results	Accuracy	Precision	Recall	F1-Score
Trial 1	98.10	97.48	98.44	97.73
Trial 2	99.60	99.53	99.55	99.60
Trial 3	98.50	98.58	98.55	98.57
Trial 4	98.50	98.60	98.92	98.75
Trial 5	97.30	97.24	97.84	97.53
Trial 6	97.70	98.04	97.71	97.84
Trial 7	98.90	98.82	98.84	98.88
Trial 8	98.50	98.68	98.33	98.34
Trial 9	98.10	98.24	98.21	98.28
Trial 10	98.90	98.75	98.93	98.76
Average	98.41	98.39	98.53	98.42

Table 4. Metrics of all ten trials.

Metrics	In percentage (%)
Top_1 accuracy	99.60
Top_5 accuracy	100
Recall	99.55
Precision	99.53
F1- Score	99.60

Table 5. Metrics of best trial.

further supports the selection of YOLOv8 for our study. These findings justify the selection of YOLOv8 in our study over newer yet less stable alternatives like YOLOv11.

Table 5 highlights the peak performance of the DL model during its most successful trial in diagnosing cotton diseases. In this best trial, the model achieved a Top_1 accuracy of 99.60%, indicating that it correctly identified the disease as its top prediction nearly every time. The Top_5 accuracy remained at 100%, ensuring that the correct diagnosis was always included within the top five predictions. The recall was 99.55%, demonstrating the exceptional ability of the model to correctly identify nearly all actual disease cases, minimizing the likelihood of missed diagnoses. With a precision of 99.53%, the model demonstrated that nearly all of its positive predictions were accurate, effectively reducing the number of false positives. F1_Score of 99.60% balanced out the precision and recall results, proving the efficiency of the model in cotton disease detection. This best trial underscores the superior performance of the model, which shows its potential as a highly reliable tool for precision agriculture (Figs. 10, 11, 12, 13 and 14).

Table 6 reveals the strength and high accuracy of the proposed model in diagnosing cotton diseases. The Top_1 accuracy of the model was equal to 98.41%, meaning that in almost all cases, the disease was predicted correctly as the top prediction. The Top_5 accuracy grew to 100%, ensuring the correct disease was always present among the first five predictions and emphasizing the reliability of the model. The model performed pretty well on the test set: 98.53% Recall, meaning it had the ability to effectively identify almost all cases of disease, which minimizes missed diagnoses; 98.39% precision, meaning most positive predictions by the model are correct, thus avoiding false positives. This makes the F1-score 98.42%, indicating that this model is highly effective and consistent over ten separate trials. These indicators support the ability of the model to accurately and reliably diagnose cotton diseases, being of use in precision agriculture.

Figure 15 is a confusion matrix presenting the validation performance of a classification model on different diseased leaves. True labels have been mapped on the x-axis, prediction labels have been mapped on the y-axis, true classifications are illustrated on the diagonal cells, and the off-diagonal cells illustrate misclassifications. This model performs the job extremely well in classification, where it is able to identify “Curl Virus” in 53 out of 53 samples, “Leaf Redding” in 72 out of 72, and “Herbicide Growth Damage” in 34 out of 34. However, out of the 263 samples, the model made just one confusion between the classes “Healthy Leaf” and “Bacterial Blight”. Figure 14 illustrates the normalised confusion matrix (best validation) (Fig. 16).

Prediction of the best trial shows the prediction of the validation set of best trial with an accuracy of 99.60%, and all 16 leaves in the above images are correctly classified. The output of the proposed approach shows the class of each cotton diseased leaf in the left corner of each image, as shown in Fig. 17.

Discussion

The methodology presented in this work employs a robust approach to classify cotton diseases using a DL model based on YOLOv8. Our results proved that DL, with the help of the YOLOv8 classification model and a 10-fold cross-validation technique, can diagnose diseases in cotton leaves for precision agriculture. The advantage of using SAR-CLD 2024 sourced from NCRI, Gazipur, has been robust testing on diverse leaf images in a real-

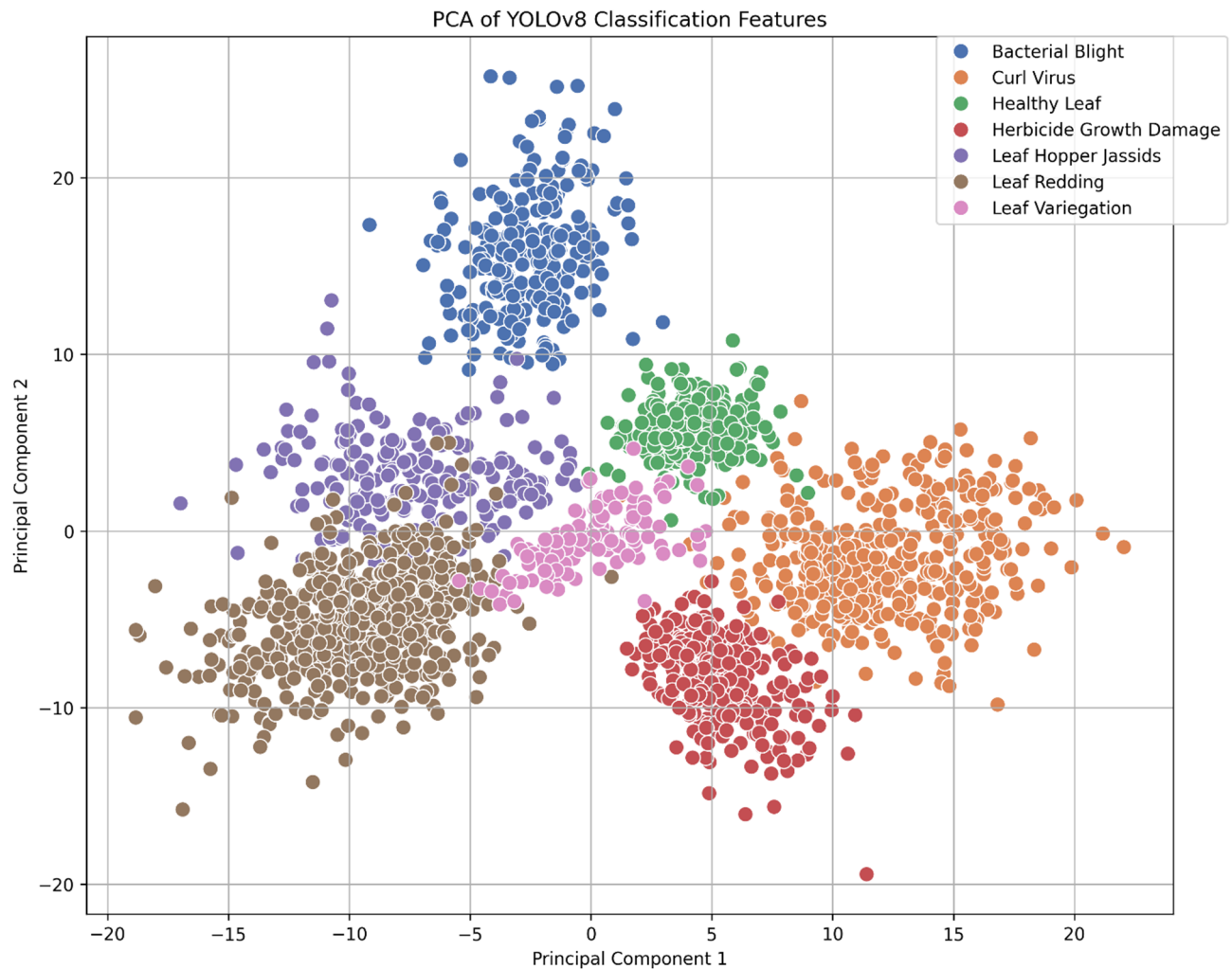


Fig. 8. Principal Component Analysis (PCA) on the deep feature vectors extracted from the final layer of the YOLOv8 classifier.

time environment. This comprehensive coverage of conditions in the dataset is essential for creating a model capable of distinguishing between various diseases and stress factors. Additionally, the dataset is thoughtfully organized into training, validation, and testing sets, ensuring that the model undergoes a thorough evaluation, which is crucial for developing a reliable disease classification tool. Cross-validation helps deal with overfitting and enhances the generalizability of the model when exposed to different subsets of datasets for both training and testing. In contrast²⁰, Elaraby et al. obtained an accuracy of 98.83% for multi-crop disease classification using the PlantVillage dataset²¹. Pan et al., in 2024, whose model CDDLite-YOLO achieved a mAP of 90.6%. Additionally²², Ahmed (2021) and²³Gao et al. (2024) employed transfer learning and YOLOv8 to further improve cotton disease detection, with a development accuracy in both cotton pest and cotton disease detection set at 94%. A key aspect of the study is the use of k-fold cross-validation, which divides the dataset into multiple folds. This technique is essential for ensuring that the model performs robustly across various data subsets. It is particularly important in agricultural applications, where environmental variations can significantly impact the appearance of cotton leaves. By utilizing k-fold cross-validation, the model is exposed to a wide range of disease symptoms and ecological conditions, which enhances its ability to generalize and reduces the risk of overfitting. When combined with advanced DL architectures like YOLOv8, this method ensures that the model can perform effectively in real-world scenarios.

The YOLOv8m-cls model used for image classification in this study demonstrated high effectiveness. Both confidence and class labels are mapped for each image, ensuring that the measure of certainty of classification is also evaluated along with the predicted class. This feature is particularly beneficial in precision agriculture, where knowing the specific class of an image is often sufficient for decision-making without the need to localize individual objects within the image. The YOLOv8 architecture, consisting of 141 layers and millions of parameters, enables fast and accurate classification, making it well-suited for large-scale deployment in field conditions³¹. Shahid et al., (2024) used GoogleNet, achieving 93.40% accuracy and 95% F1_score, AlexNet achieved accuracy 93.40%, and InceptionV3 achieving accuracy 91.80%²⁹. Rai and Pahuja (2023) used DCNN to achieve 97.98% accuracy²⁵Li et al., (2024) used CFNet-VoV-GCSP-LSKNet-YOLOv8s achieving 89.9%

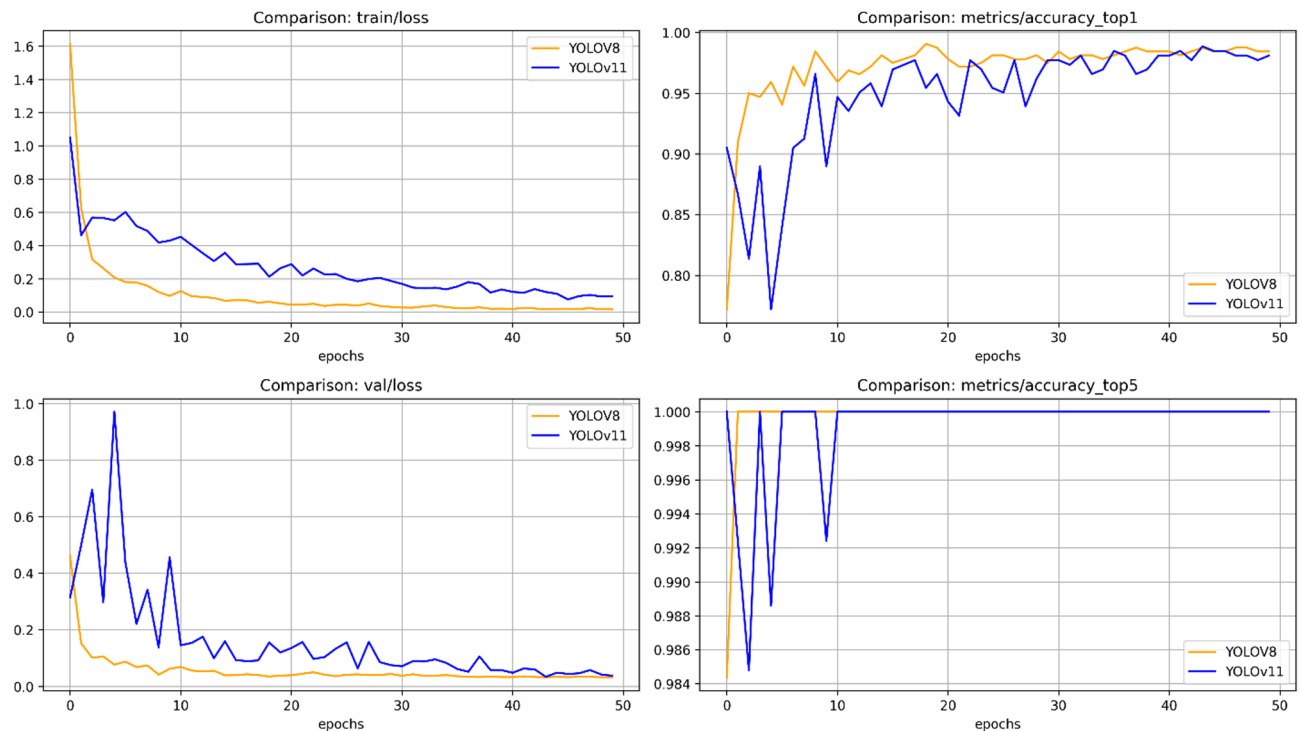


Fig. 9. Comparative analysis between YOLOv8 and YOLOv11.

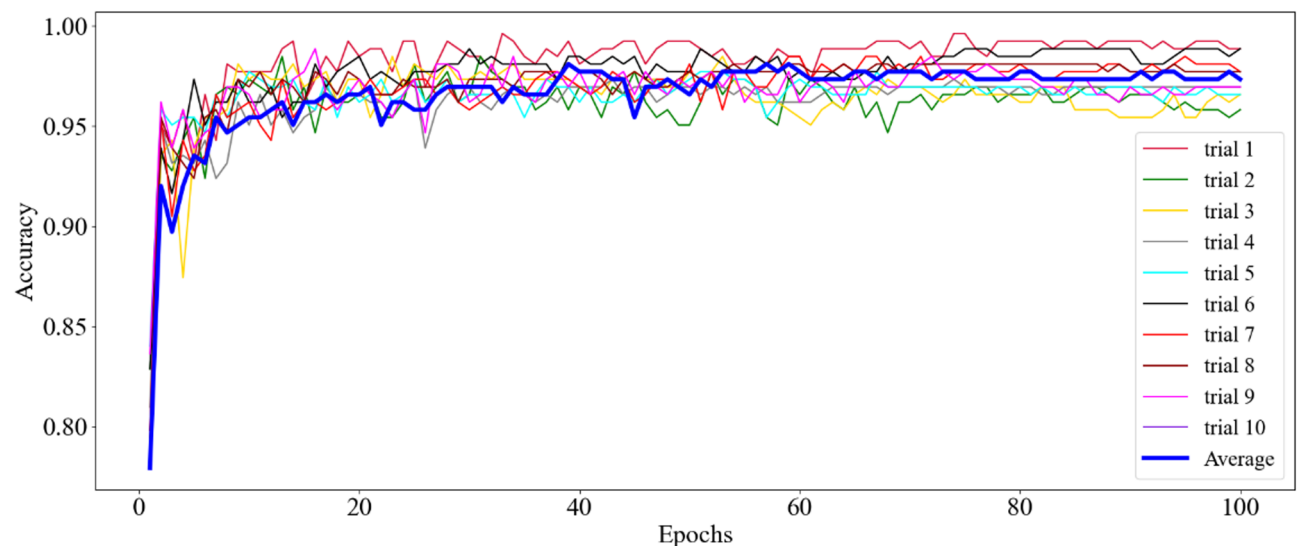


Fig. 10. Accuracy vs. epochs of each trial: shows the accuracy Vs epochs for each trial, and the dark blue line shows the average of all ten trials.

precision²⁶. Nazeer et al., (2024) identified curl disease with 99% accuracy. This study deals only with detecting Cotton Leaf Curl Disease. Many current datasets, such as those used by²⁷Kolachi et al. (2023) and⁸Latif et al. (2021), are limited by the number of classes or environmental conditions they capture. While their model was effective for a specific application, the proposed model in this study surpasses these results by achieving a higher degree of accuracy in a more complex task, as the proposed model has seven classes in the dataset.

The experimental results highlight the effectiveness of the proposed approach. The model achieved Top_1 and Top_5 accuracy of 99.60% and 100% respectively. Top_1 accuracy demonstrates accurate detection in the first attempt and Top_5 accuracy demonstrated overall accuracy. Minimisation of F.P. has been ensured by 99.55% recall and 99.53% precision results. These metrics, along with an F1 score of 99.60%, underscore the exceptional performance and robustness of the model. The key aspect of the study is the utilization of 10-fold

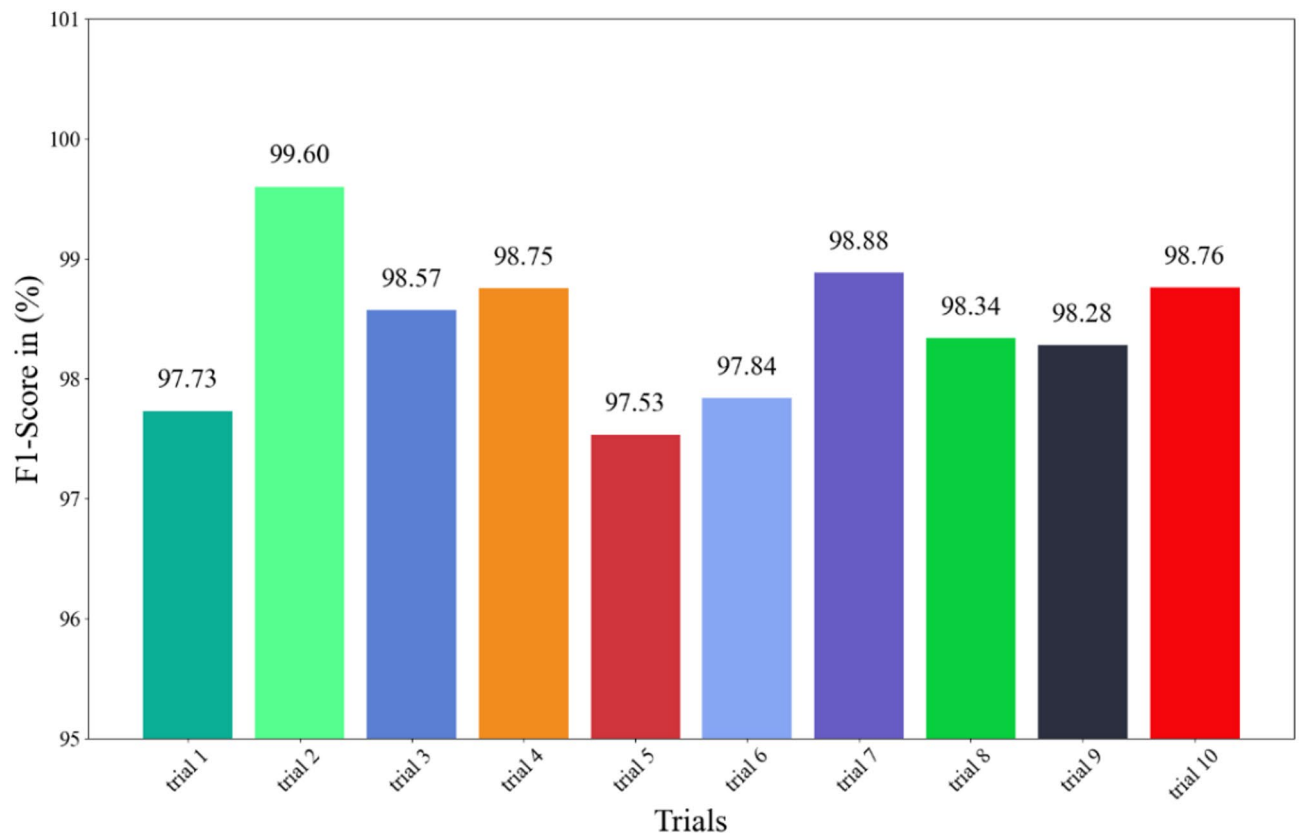


Fig. 11. Graph of F1-score for each trial: shows the bar chart of the F1-score in each trial. The best trial was observed to be trial 2.

cross-validation, which offers a more robust performance than single train-test splits. By rotating through the dataset and using every sample as part of the training and validation set, the model was able to avoid overfitting, a common challenge in DL models for agriculture due to limited or skewed datasets. The k-fold approach, as shown by the consistent average Top_1 accuracy of 98.41% and recall of 98.53% across all trials, provided more robust and generalizable results.

Table 6 illustrates consistently high Top_1 and Top_5 accuracy, of 98.41% and 100% respectively, across 10 trials. The values for precision, recall, and F1-score further support the reliability of the model, making it a promising tool for diagnosing cotton diseases in practical applications. The confusion matrix (Fig. 16) also highlights the excellent classification ability of the model, with very few misclassifications. This indicates that the model can reliably diagnose diseases such as “curl virus,” “leaf redding,” and “herbicide growth damage” with minimal error.

This study has been able to achieve an accurate and reliable model for cotton disease detection, which outperformed the majority of contemporary models when tested on a diverse range of metrics. 10-fold cross-validation integration ensured robustness of the model for real-time usage.

Conclusion

This study proposed an efficient YOLOv8 classification model integrated with 10-fold cross-validation for improving the robustness and scalability of the model. This method has been able to outperform with 99.60% Top_1 accuracy, and 100% Top_5 accuracy. The method exhibited a high precision, recall, and F1-score level, which showed an accurate and robust approach to diagnosing multiple diseases on cotton leaves. The model could effectively use k-fold cross-validation to minimize overfitting, hence performing very well over different data subsets, a feature critical to practical agricultural systems. The proposed model exceeded the benchmark accuracies and remedies some limitations noted from available literature: low classes, controlled datasets, and inadequacy with adaptability to field conditions. This evidently showed its potential to be used as a very important tool in precision agriculture that will give timely disease detection with great accuracy, thus reducing crop losses while improving cotton yield. Future work will include more data collection from the field in real time and include environmental variables that might affect detection.

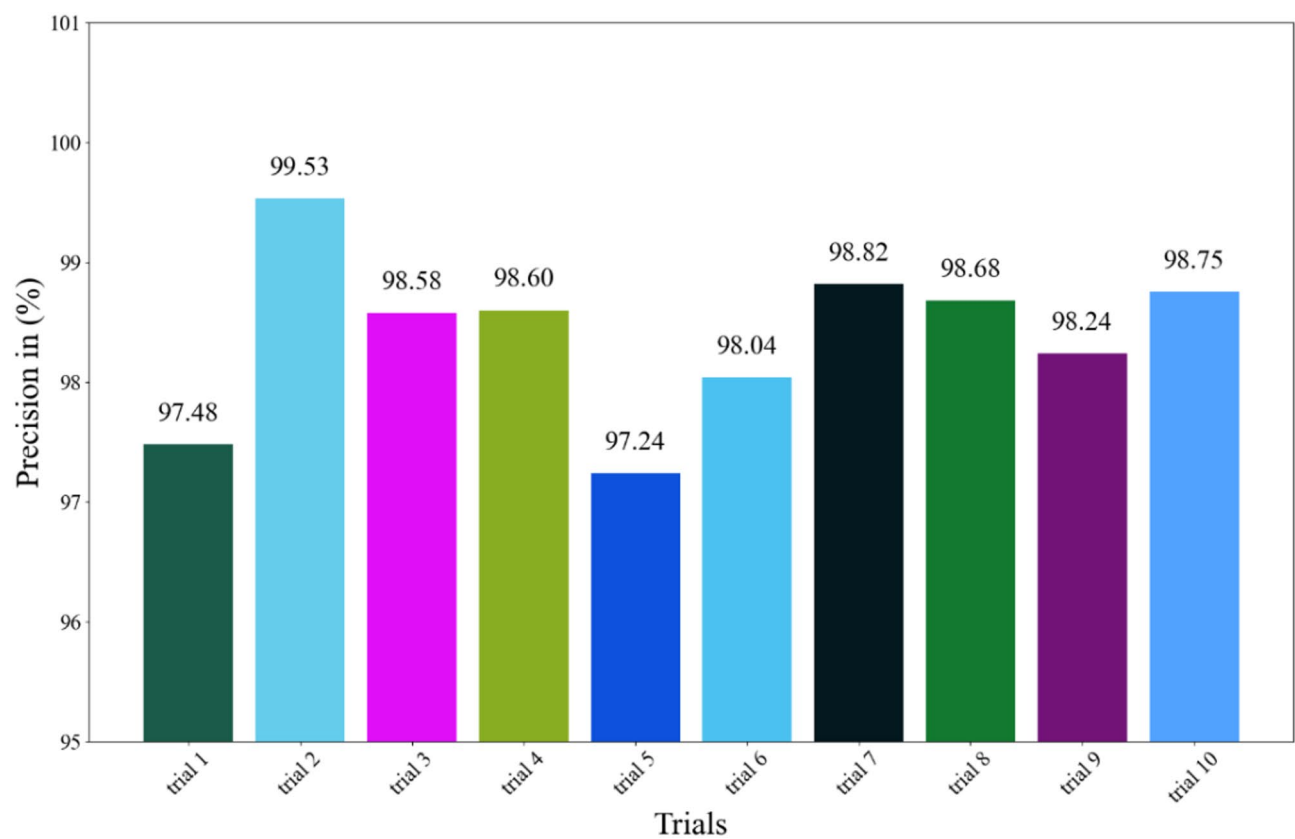


Fig. 12. Graph of precision for each trial: shows the bar graph of the precision of each trial. Trial two performed the best in all the trials.

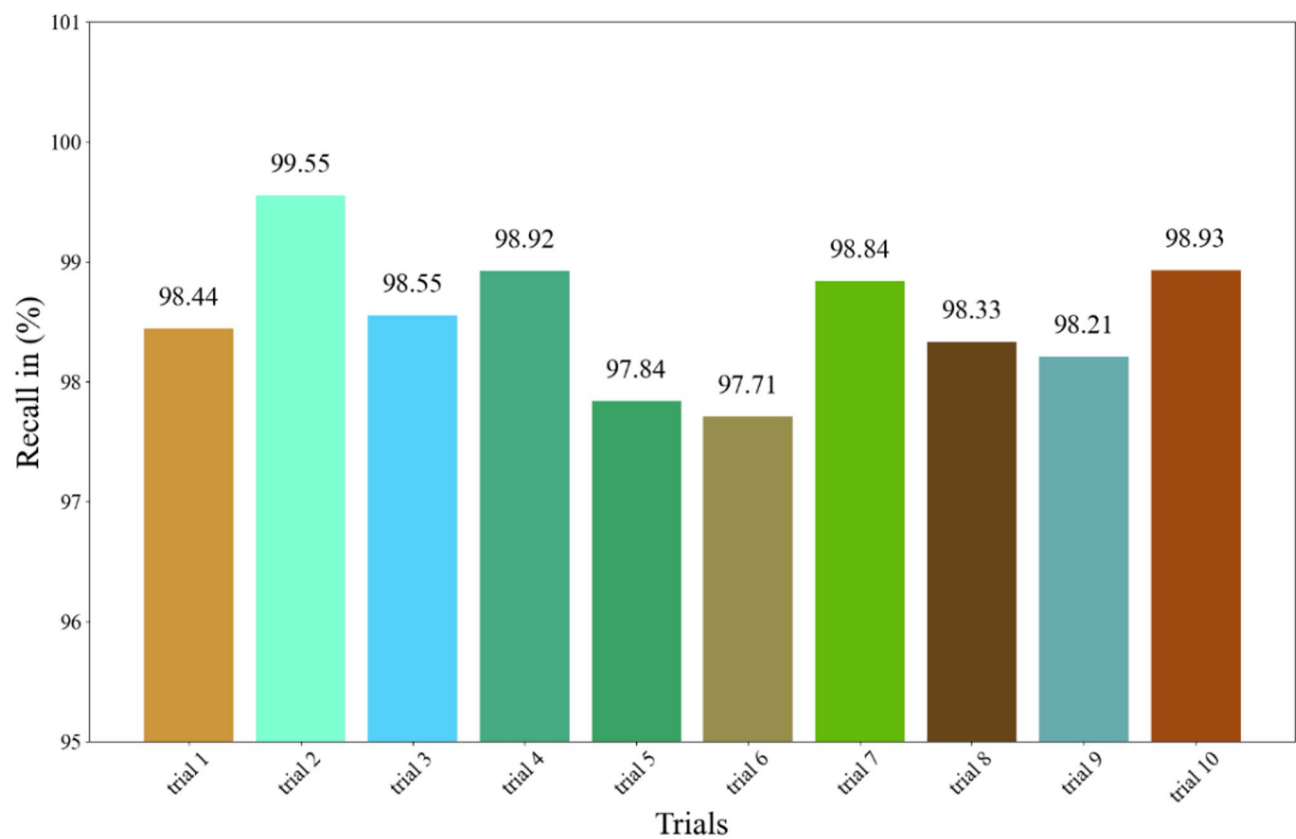


Fig. 13. Graph of Recall for each trial: shows the bar chart of the recall of each trial. Trial two showed excellent results of 99.55% in all ten trials.

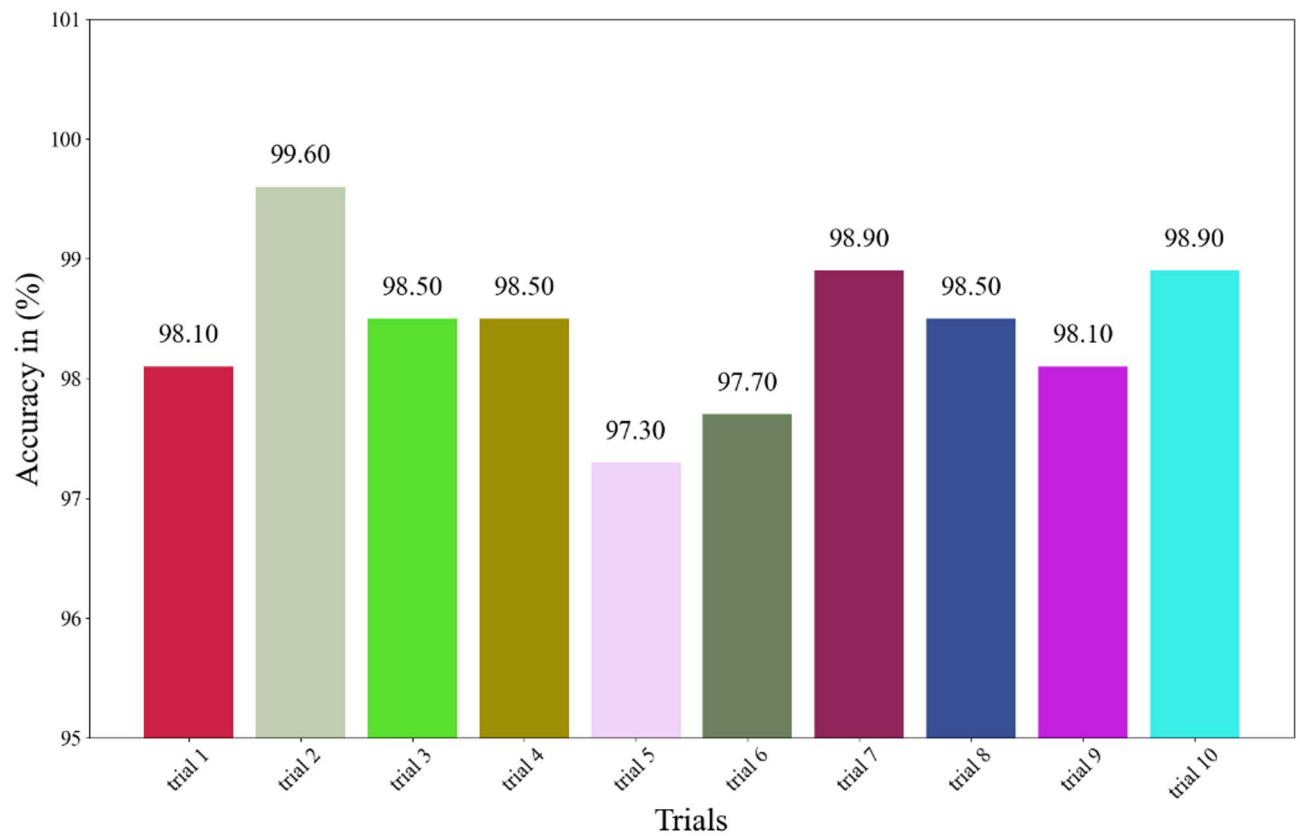


Fig. 14. Graph of Accuracy for each trial: shows the bar chart of the accuracy of each trial. The best trial was trial two, which showed an outstanding accuracy of 99.60%.

Metrics	In percentages (%)
Top_1 accuracy	98.41
Top_5 accuracy	100
Recall	98.53
Precision	98.39
F1-Score	98.42

Table 6. Shows the average of all five metrics of ten trials. The average accuracy of ten trials is noted as 98.41%.

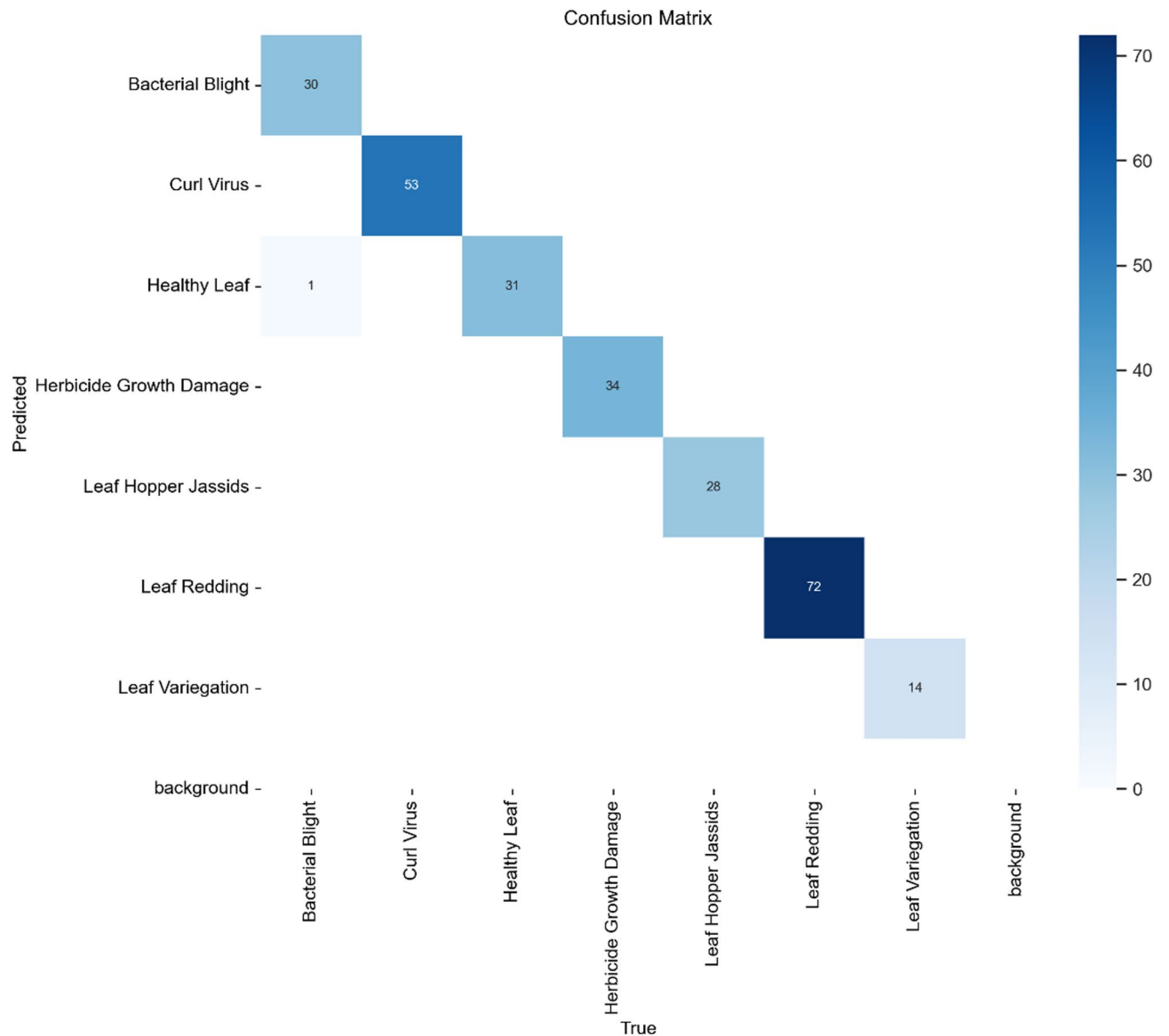


Fig. 15. Confusion matrix of best trial validation shows the confusion matrix of the validation dataset done by the best trial. In this confusion matrix, all the images are correctly classified as the labels given to them, and only a single image was not correctly identified as the true value. Out of 263 images, only one image was not predicted correctly; otherwise, all the predictions were correct. The accuracy of the given matrix is 99.60%.

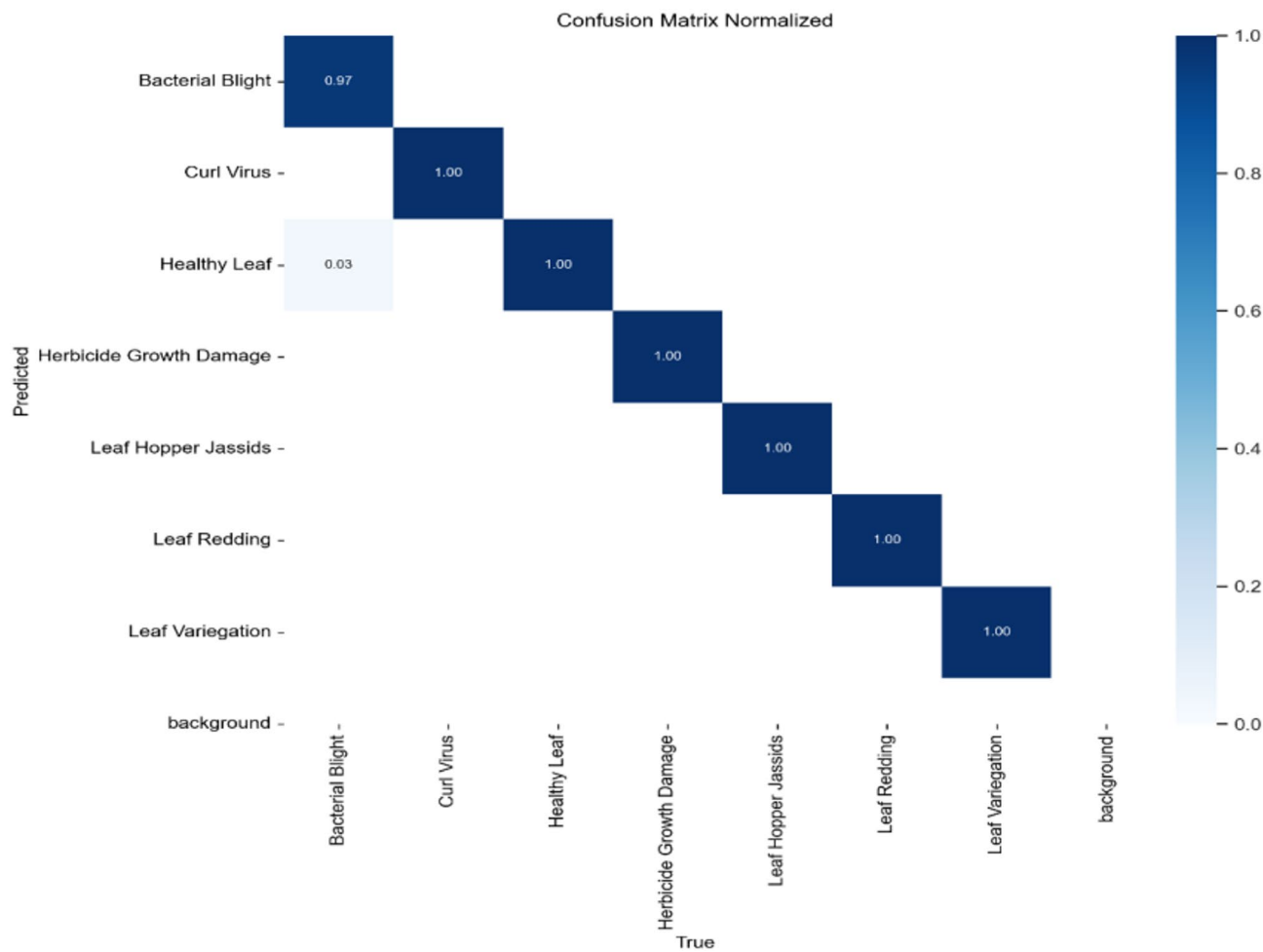


Fig. 16. Shows the normalized confusion matrix of the best trial. All the predictions are clearly done correctly, and only one image is not detected correctly.



Fig. 17. Prediction of the best trial: shows the prediction of the validation set of the best trial with an accuracy of 99.60%, and all 16 leaves in the above images are correctly classified. The output of the proposed approach shows the class of each cotton diseased leaf in the left corner of each image.

Data availability

The data is available at “<https://doi.org/10.17632/b3jy2p6k8w.2> <https://data.mendeley.com/datasets/b3jy2p6k8w/2>”.

Received: 3 April 2025; Accepted: 22 July 2025

Published online: 13 October 2025

References

1. Padhiary, M., Saha, D., Kumar, R., Sethi, L. N. & Kumar, A. Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation. *Smart Agricultural Technol.* **8**, 100483 (2024).
2. Getahun, S., Kefale, H. & Gelaye, Y. Application of Precision Agriculture Technologies for Sustainable Crop Production and Environmental Sustainability: A Systematic Review. *The Scientific World Journal* (2024). (2024).
3. Sanyaolu, M. & Sadowski, A. The role of precision agriculture technologies in enhancing sustainable agriculture. *Sustainability* **16**, 6668 (2024).

4. Khan, S. U. et al. A review on automated plant disease detection: motivation, limitations, challenges, and recent advancements for future research. *Journal King Saud Univ. Comput. Inform. Sciences* **37**, (2025).
5. Jung, M. et al. Construction of deep learning-based disease detection model in plants. *Sci. Rep.* **13**, 7331 (2023).
6. Nautiyal, S., Dimri, S., Riyal, L., Sharma, H. & Dwivedi, C. Natural fibres and their composites: a review of chemical composition, properties, Retting methods, and industrial applications. *Cellulose* <https://doi.org/10.1007/s10570-025-06487-x> (2025).
7. Khan, M. A. et al. Impacts of climate change on cotton production and advancements in genomic approaches for stress resilience enhancement. *Journal Cotton Research* **8**, (2025).
8. Rizwan Latif, M. et al. Cotton leaf diseases recognition using deep learning and genetic algorithm. *Computers Mater. Continua.* **69**, 2917–2932 (2021).
9. Thivya Lakshmi, R. T. Visu. CoDet: A novel deep learning pipeline for cotton plant detection and disease identification. *Automatika* **65**, 662–674 (2024).
10. Hirenkumar Kukadiya, Arora, N., Srivastava, S. & Divyakant Meva & An ensemble deep learning model for automatic classification of cotton leaves diseases. *Indonesian J. Electr. Eng. Comput. Sci.* **33**, 1942–1942 (2024).
11. Singla, A. et al. Exploration of machine learning approaches for automated crop disease detection. *Curr. Plant. Biology.* 100382–100382. <https://doi.org/10.1016/j.cpb.2024.100382> (2024).
12. Joshi, K. et al. Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation. *Curr. Plant. Biology.* 100437–100437. <https://doi.org/10.1016/j.cpb.2025.100437> (2025).
13. Jafar, A., Bibi, N., Naqvi, R. A. & Jeong, D. Abolghasem Sadeghi-Niaraki Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. *Frontiers Plant. Science* **15**, (2024).
14. Minhans, K., Sharma, S., Sheikh, I., Alhewairini, S. S. & Sayyed, R. Artificial intelligence and plant disease management: an Agro-Innovative approach. *Journal Phytopathology* **173**, (2025).
15. Terven, J., Córdova-Esparza, D. M. & Romero-González, J. A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **5**, 1680–1716 (2023).
16. Luo, J. et al. Efficient small object detection you only look once: A small object detection algorithm for aerial images. *Sensors* **24**, 7067 (2024).
17. Subeesh, A. & Mehta, C. R. Automation and digitization of agriculture using artificial intelligence and internet of things. *Artif. Intell. Agric.* **5**, 278–291 (2021).
18. Ayoub Shaikh, T., Rasool, T. & Rasheed Lone, F. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **198**, 107119 (2022).
19. Md, M. et al. A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture. *Intell. Syst. Appl.* **20**, 200278–200278 (2023).
20. Elaraby, A., Hamdy, W. & Alruwaili, M. Optimization of deep learning model for plant disease detection using particle swarm optimizer. *Computers Mater. Continua.* **71**, 4019–4031 (2022).
21. Pan, P. et al. Lightweight cotton diseases real-time detection model for resource-constrained devices in natural environments. *Frontiers Plant. Science* **15**, (2024).
22. Ahmed, M. R. Leveraging convolutional neural network and transfer learning for cotton plant and leaf disease recognition. *Int. J. Image Graphics Signal. Process.* **13**, 47–62 (2021).
23. Gao, R. et al. Intelligent cotton pest and disease detection: edge computing solutions with transformer technology and knowledge graphs. *Agriculture* **14**, 247–247 (2024).
24. Bharathi, S. L., Deepa, N., Sathya, J., Priya & Muthulakshmi, K. Innovative agricultural diagnosis: DQRR-AFH algorithm model for effective leaf disease prevention and monitoring. *Earth Sci. Inf.* **17**, 2461–2476 (2024).
25. Li, R. et al. Identification of cotton pest and disease based on CnFNet- VoV-GCSP -LSKNet-YOLOv8s: a new era of precision agriculture. *Frontiers Plant. Science* **15**, (2024).
26. Nazeer, R. et al. Detection of cotton leaf curl disease's susceptibility scale level based on deep learning. *Journal Cloud Computing* **13**, (2024).
27. Kolachi, A. R., Soomro, S. R., Baloch, S. K., Patoli, A. A. & Anwar, S. Cotton leaf disease classification using YOLO deep learning framework and Indigenous dataset. *Int. J. Sys Innov.* **7** (7), 80–88. [https://doi.org/10.6977/IJoSI.202309_7\(7\).0005](https://doi.org/10.6977/IJoSI.202309_7(7).0005) (2023).
28. Zhu, D., Feng, Q., Zhang, J. & Yang, W. Cotton disease identification method based on pruning. *Frontiers Plant. Science* **13**, (2022).
29. Rai, C. K. & Pahuja, R. Classification of diseased cotton leaves and plants using improved deep convolutional neural network. *Multimedia Tools Appl.* <https://doi.org/10.1007/s11042-023-14933-w> (2023).
30. Chitraranjan Kumar Rai & Roop Pahuja. An ensemble transfer learning-based deep Convolution neural network for the detection and classification of diseased cotton leaves and plants. *Multimedia Tools Appl.* <https://doi.org/10.1007/s11042-024-18963-w> (2024).
31. Shahid, M. F. et al. An ensemble deep learning models approach using image analysis for cotton crop classification in AI-enabled smart agriculture. *Plant. Methods.* **20**, 104 (2024).
32. Bishhash, P., Nirob, M. A. S., Shikder, M. H., Sarower & Afjal SAR-CLD-2024: A comprehensive dataset for cotton leaf disease detection. *Mendeley Data.* **V2** <https://doi.org/10.17632/b3jy2p6k8w.2> (2024).
33. Anwar, S., Soomro, S. R., Baloch, S. K., Patoli, A. A. & Kolachi, A. R. Performance analysis of deep transfer learning models for the automated detection of cotton plant diseases. *Eng. Technol. Appl. Sci. Res.* **13**, 11561–11567 (2023).
34. Chepuri, S. & Ramadevi, Y. A novel fusion study on disease detection in cotton plants using embedded approaches of neural networks. *Lecture Notes Networks Syst.* 171–181. https://doi.org/10.1007/978-981-99-9704-6_15 (2024).
35. Jai Vignesh, P. S., Adhish, K., Rithik, R., Sanjeev, R. & Rajesh, C. B. S. Model Validation to Enhance Precision Agriculture Using DeepDream and Gradient Mapping Techniques. *Lecture Notes in Networks and Systems* 359–372 (2022). https://doi.org/10.1007/978-981-19-4960-9_28
36. Gayatri, N., Vamsi, B., Vidyullatha, P., Deep Learning, L. S. T. M. & Approach on Hyperspectral Images using Keras Framework. *International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (2022) (2022) (2022). <https://doi.org/10.1109/icscds53736.2022.9760833>
37. Kumar, M., Arora, A., Deb, A. & Yadav, A. L. Deep Learning for Accurate Plant Disease Classification Using ResNet50: A Comprehensive Approach. *International Conference on Computational Intelligence and Computing Applications (ICCICA)* 125–130 (2024) 125–130 (2024) (2024). <https://doi.org/10.1109/ICCICA60014.2024.10584860>

Acknowledgements

Work on plant abiotic stress tolerance in SSG laboratory was partially supported by the University Grants Commission (UGC), Science and Engineering Research Board (SERB), Council of Scientific & Industrial Research (CSIR), Govt. of India. RG, SSG also acknowledges partial support from DBT-BUILDER grant (No. BT/INF/22/SP43043/2021). We sincerely apologize to our contemporaries whose work has not been discussed in this article due to space restrictions.

Author contributions

KJ, NT, RN, KS, SSG, RG conceptualized the concept; KJ, YY, SH, BS, RG performed the research; KJ, YY, SH,

BS, AN, RG analysed the research and wrote the manuscript; KJ, RN, NT, BS, KS, SSG, RG read and edited the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.S., R.G. or S.S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025