



Article

An Improved Detection Method for Crop & Fruit Leaf Disease under Real-Field Conditions

Serosh Karim Noon^{1,*}, Muhammad Amjad^{2,†}, Muhammad Ali Qureshi^{2,†}, Abdul Mannan^{1,†} and Tehreem Awan^{1,†}

¹ Department of Electrical Engineering, NFC Institute of Engineering & Technology, Multan 59060, Pakistan; mannan@nfciet.edu.pk (A.M.); tehreemawan@nfciet.edu.pk (T.A.)

² Department of Electronic Engineering, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan; muhammad.amjad@iub.edu.pk (M.A.); ali.qureshi@iub.edu.pk (M.A.Q.)

* Correspondence: seroshkarim@nfciet.edu.pk

† These authors contributed equally to this work.

Abstract: Using deep learning-based tools in the field of agriculture for the automatic detection of plant leaf diseases has been in place for many years. However, optimizing their use in the specific background of the agriculture field, in the presence of other leaves and the soil, is still an open challenge. This work presents a deep learning model based on YOLOv6s that incorporates (1) Gaussian error linear unit in the backbone, (2) efficient channel attention in the basic RepBlock, and (3) SCYLLA-Intersection Over Union (SIU) loss function to improve the detection accuracy of the base model in real-field background conditions. Experiments were carried out on a self-collected dataset containing 3305 real-field images of cotton, wheat, and mango (healthy and diseased) leaves. The results show that the proposed model outperformed many state-of-the-art and recent models, including the base YOLOv6s, in terms of detection accuracy. It was also found that this improvement was achieved without any significant increase in the computational cost. Hence, the proposed model stood out as an effective technique to detect plant leaf diseases in real-field conditions without any increased computational burden.

Keywords: deep networks; plant disease; agriculture; image processing; YOLOv6



Citation: Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A.; Awan, T. An Improved Detection Method for Crop & Fruit Leaf Disease under Real-Field Conditions.

AgriEngineering **2024**, *6*, 344–360.
<https://doi.org/10.3390/agriengineering6010021>

Academic Editor: Murali Krishna Gumma

Received: 29 December 2023

Revised: 4 February 2024

Accepted: 6 February 2024

Published: 9 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Crops, fruits, and vegetables are the economic backbone of a country. Pakistan's GDP relies on some major cash crops for domestic use & it also got its hefty share from exports of seasonal fruits [1]. Outdated agri-practices and inefficient use of technology have drastically affected local utilization and foreign trade. Pathogenic infestation hinders growth and consequently reduces the yield of the agri-products [2].

Wheat and cotton are Pakistan's most important cash crops, which are extensively grown in the Punjab province. However, the yield of the wheat crops has been adversely affected by various factors in the last few years [3]. The environmental conditions of Sindh and Punjab favor the growth of fungal pathogen (*Puccinia* & *Urocystis tritici*) that are considered to be more destructive agents for wheat, resulting in the occurrence of rust & smut that affect all parts of the plant and stunts its growth [4,5]. The country is also the fourth-largest producer of mangoes in the world [6]. Anthracnose is considered the most devastating disease for the growth and yield of mango. It starts from leaf and twigs and spreads in all parts of the mango plant [7].

Timely remedial actions can prevent the spread of the disease. Over the past few years, the low production of these crops and fruits has led agriculturists to adopt modern agricultural methods [8]. One such method is the timely identification of disease on the plant and the recommendation of an effective treatment for it. Farmers depend on visual

symptoms in conventional farming to identify the specific disease type and its stage. However, visual symptoms can often be similar for multiple diseases and can be influenced by weather conditions. Accurately distinguishing between biotic pathogen attacks and abiotic nutritional deficiencies requires a considerable level of technical expertise or assistance from pathologists [9].

For that reason, machine learning and deep learning techniques are extensively being used in stress and disease identification of plants due to their ability to learn deep intricate features to be learned from image datasets at better speed and accuracy [10]. With the increase in sustainable agriculture the need to employ computer vision-based techniques to revolutionize. Object detection-based methods have been improvised in the past few years to improve weed identification, pest control, and plant disease detection [11]. One of the challenges faced by the researchers is to locate and classify the type of stress in real-world scenarios [12]. The more robust and accurate the model will be and trained on large datasets that can be used for agricultural applications. Such trained models can be deployed on hardware platforms or embedded systems to develop automatic plant disease detection platforms [13]. Convolutional neural network-based techniques [10,14] are extensively used to extract relevant information from the diseased area. Unlike the traditional machine learning-based methods [15], it does not classify the diseased area on the lesion's colour, background, size, and shape. The same cannot be used for large datasets or real-field agricultural applications.

Recent studies have extensively used computer vision-based object detection techniques as an automated tool to locate and classify the plant stress type [9,16]. To localize and classify the diseased part in an image, researchers worldwide have thrived to look for an accurate and efficient model. Saleem et al. [17] have trained & fine-tuned different meta architectures like SSD, RFCN, and Faster RCNN to detect 26 diseased and 12 healthy parts of the plants. A training accuracy of 73.07% was obtained for the SSD model while using the Adam optimizer. Authors in [18] proposed a novel SSD-based model fusing attention mechanism with VGG feature extractor. An improved accuracy performance was observed for the PlantVillage dataset. In another work [19], authors have localized and classified diseased areas of plants using a novel deep-learning framework. The improved RefineDet model achieves a remarkable accuracy of 99.994% on the plantVillage dataset. The quest to attain better accuracy at low cost has been the prime objective of researchers in the past few years. Chowdhury et al. [20] analyzed the performance of lightweight models on plain and segmented images of tomato leaves. The EfficientnetB4 model achieved 99.89% accuracy in classifying 10 tomato healthy and diseased classes. Wang et al. [21] proposed a lightweight YOLOv5 model to attain improved detection results on public and self-collected datasets. The model weight reduction was achieved using GhostNet and the weighted box fusion method. Most of the work showcased near-to-ideal accuracies on datasets having images with plain or clear backgrounds or a single leaf image. The detection ability of these trained models will be limited when they are tested on images captured in varying and difficult environmental conditions. A novel Bidirectional transposition feature pyramid network [22] was proposed to detect apple leaf diseases in complex real-field conditions with a remarkable accuracy performance. The cross-attention module was used to detect relevant feature information. In another notable work, Zhao et al. [23] integrated a coordinate attention module into the backbone of the You Only Look Once (YOLOv5s) model to detect small buds and occluded flowers in real-field conditions effectively.

Although the research in the field of deep learning-based disease detection in plants is quite mature at the moment, locating and identifying the type of stress in real-field conditions remains challenging for computer vision experts in the field. The main aim of the proposed work is to extract diseased areas on the respective plant considering the unconstrained environmental conditions. Images, when captured in real-field conditions, may vary in scale because of the inconsistent distance between the target leaf/plant and the camera. The diseased plant in the background of the target leaf can also be the cause of scale invariance. Moreover, varying lighting conditions, complex & similar backgrounds,

and variability of diseased symptoms pose further challenges for the detection model. A few images, shown in Figure 1 from the self-collected dataset, explain these typical scenarios.

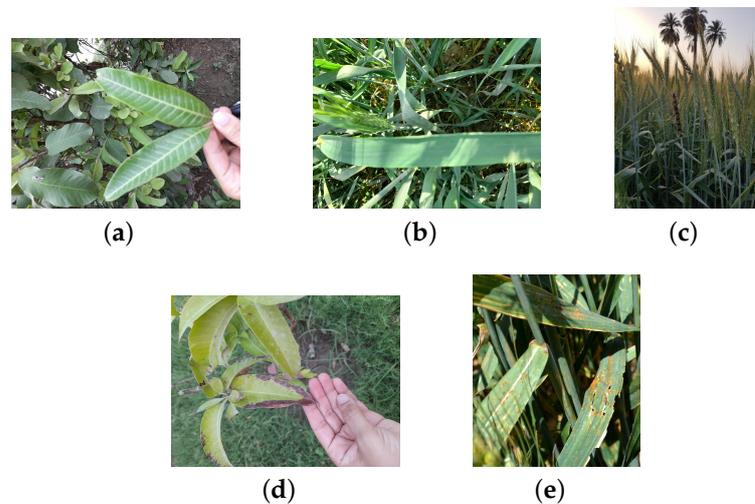


Figure 1. Sample images taken from the dataset showing difficult field conditions (a) Similar background (b) Shadow interference (c) varying light & complex background (d) variability of diseased symptoms (e) multiple objects in varying light.

In this paper, we propose an improved deep-learning framework to localize and classify various plant diseases in real-field conditions. Major contributions of the proposed model are given as follows:

- A fine-tuned efficient model (based on YOLOv6s) was trained and optimized using Gaussian error linear unit (GELU) in the backbone of the model. That improved model's generalization in detecting small and complex objects.
- Efficient channel attention was introduced in the basic Rep Block in the neck region of the base model (YOLOv6s) to improve the accuracy and recall of the detection model without any additional computational cost.
- To improve the regression accuracy, the Generalized-IoU (GIoU) loss in the base YOLOv6s model is replaced with the SCYLLA-IoU (SIoU) loss function in the proposed model.
- The authors present a self-collected dataset comprising 3305 images captured in real field conditions.

2. Materials and Methods

2.1. Object Detection

When it comes to classifying and localizing a particular class in a complex scene, object detection algorithms play a vital role [18]. The detection performance of deep learning-based object detectors has gained breakthrough performance in computer vision [24]. Object detection models employ deep learning methods to form single-stage and two-stage detectors. Localization and classification of the object are carried out simultaneously in two-stage detectors so they offer superior performance [25]. However, SSD models directly classify and form a bounding box around the object, thus making the detection process faster, but accuracy may be compromised. Most recent studies have employed single-stage detectors over two-stage detectors by working on the methods to improve the classification and detection accuracy [18,26,27]. Also, the use of anchor-free detectors [28] made the inference process simpler and more generalized compared to anchor-based methods. The study focuses on a swifter and accurate detecting model that can be employed for real-time detection; hence YOLOv6 model was employed for the detection of plant disease for our specific dataset due to its better accuracy performance as compared to other detectors with similar inference speed [29].

2.2. YOLOv6 Model

YOLOv6 [28] is a deep learning-based one-stage object detector. The model is chosen for its improved baseline performance compared with more recent state-of-the-art object detection models on our specific dataset. The model outperforms others regarding inference speed, model convergence, and accuracy. The backbone, neck, and head are the basic parts of the YOLOv6 model. The reparametrized VGG-style backbone and anchor-free detection suit several hardware-based real-time applications. The anchor point-based paradigm makes it suitable to predict the detection results as the regression branch calculates the distance from all sides of the bounding box to the anchor box. The model employs Varifocal loss (VFL) [30] and Distribution Focal Loss (DFL) for detection.

The YOLOv6 model comes in several versions like YOLOv6-L, YOLOv6-M, YOLOv6-S and YOLOv6-Nano. However, the authors selected YOLOv6-S due to its reasonable accuracy and low computation cost. The feature extraction task is carried out in the backbone to the neck structure to aggregate low-level and high-level semantic information. The reparameterized backbone and neck incorporate VGG networks and skip connections. The resulting RepBlock [31] encompasses the effective classification performance of VGG and better accuracy of ResNets. The backbone and neck structures are GPU-friendly, and the model can also be used for hardware applications.

YOLOv6-s uses EfficientRep and RepBiFPAN as backbone and neck structures. Multi-scale features from reparameterized blocks are aggregated using PAN structure in the neck. The BiFusion block makes the low-level feature concatenation more effective. The aggregated features are passed to the Efficient Decoupled head, which performs the classification and regression tasks separately. This decoupled head strategy reduces the complexity and enhances the accuracy as in the YOLOX model [9]. The VFL is used as classification loss along with any IoU (Intersection Over Union) loss for regression purposes. Additionally, DFLs are used to improve bounding box localization in YOLOv6 large and medium models [28]. Unlike YOLOX, the model uses Task Aligned Assignment (TAL) as a label assignment technique instead of SimOTA. The latter is considered to be a slower technique used with anchor-free detectors.

2.3. Efficient Channel Attention

The attention mechanism enhances the focus on important features in image processing applications [27]. Recently, using SE nets has shown great performance in improving the accuracy performance of various models. Channel attention like SEnet has been employed to improve performance in image classification and segmentation [32]. The spatial dimensions of the input image are squeezed into channel-wise information to attain better accuracy performance. Attention mechanisms can be attained by aggregating important features or by combining channel and spatial attention [33].

However, all such methods come with loads of computational costs. Efficient Channel Attention (ECA) performs better as the method integrates cross-channel interaction. Adaptive kernel strategy is adopted to capture cross channel and its neighbours' information [34]. After the implementation of channel-wise global pooling, k channel is adaptively determined to perform 1D convolution. Sigmoid activation is applied afterward to generate the attention channel weights.

In our proposed model, we have added the ECA attention mechanism in Rep Block to form a RepEA Block discussed in Section 2.4.1. The computational process of Efficient channel attention [34] for an input feature map having dimensions of $H \times W \times C$ where H is the height, W is the width, and C is the number of channels. The average value for each channel is calculated as shown in Equation (1).

$$Avg(A_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W A_{cij} \quad (1)$$

The formulation for all channels to share the same parameters can be shown in Equation (2) where local cross-channel interaction is implemented with 1D Convolution using k channel.

$$G_{ECA}(A_c, \theta) = \sigma(1DConv(Avg(A_c))) \tag{2}$$

The obtained channel weight factor $G_{ECA}(A_c, \theta)$ after applying sigmoid σ to the mapped values. The weighted output feature map is obtained by element-wise multiplication of the original input feature map A_{cij} with obtained channel weight as shown in Equation (3).

$$Y_{ECA} = G_{ECA}(A_c, \theta) \odot A_c \tag{3}$$

where \odot shows element-wise multiplication. The overall structure of the ECA net is shown in Figure 2.

2.4. The Proposed Methodology

An improved YOLOv6s model is proposed with *RepEA* block used in the modified neck structure of the model discussed in Section 2.4.1. The incorporation of GELU activation in *RepBlock* to promote the model-improved performance for non-linear fitting instead of ReLU is discussed in Section 2.4.2. The proposed model is fine-tuned for training on our self-collected dataset given in Section 3.

2.4.1. RepEA Block

The main building block of YOLOv6 is the Reparametrized block, namely *RepBlock* [35]. In the small model, the *EfficientRep* backbone consists of this *RepBlocks*. *RepBlock* is a stack of *RepVGGblock*. The reparametrized VGG style blocks are a parallel addition of 3×3 convolution, a 1×1 convolution, and a batch normalization operation (BN). The results are aggregated and pass through a non-linearity operation of ReLU [35]. However, in our modified *RepEA* block, GELU activation achieves this non-linearity. Channel attention with minimum complexity by adding an ECA layer in the *Repblock*. It generated channel attention using 1D convolution as shown in Figure 2. The REPEA block is modified *RepBlock* in the neck region of YOLOv6s.

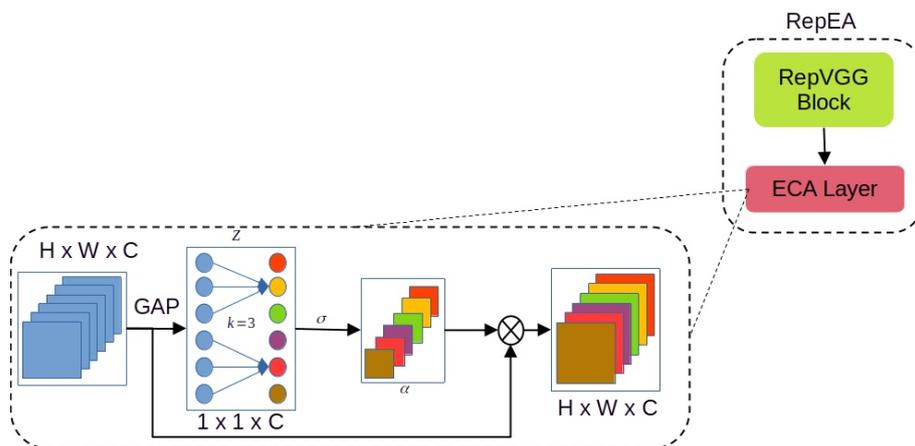


Figure 2. The structure of the RepEA block with Efficient Channel attention embedded in the Rep block of YOLOV6.

2.4.2. Gaussian Error Linear Unit (GELU)

The nonlinear activation function enables the model to learn intricate input features and establish a meaningful transformation between input and output data [36]. GELU non-linearity [37] as expressed in Equation (5) is a smooth and differentiable alternative to the ReLU function, as shown in Equation (4). It offers more smoothness than ReLU as it weights inputs based on percentile instead of sign. Therefore, GELU is popular in vision transformers and NLP AI models.

$$\text{ReLU}(x) = \max(x, 0) \quad (4)$$

$$\text{GELU}(x) = 0.5x \left(1 + \tanh \left(\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right) \right) \quad (5)$$

ReLU activation effectively imparts non-linearity but is non-differentiable at zero, leading to the problem of vanishing gradients. The proposed model uses gradient-based optimization, and GELU computes and ensures the existence of the necessary gradient for backpropagation. GELU exhibits linearity for values less than zero, as shown in Figure 3.

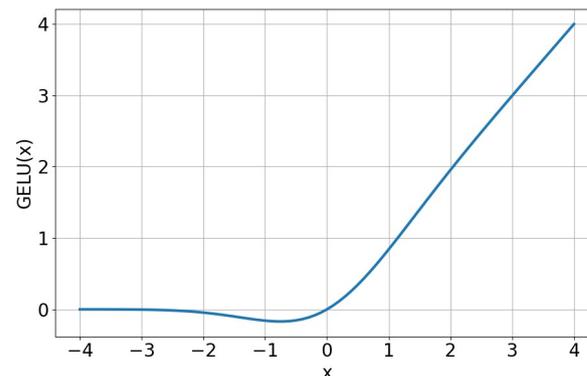


Figure 3. The GELU function used in place of ReLU.

2.4.3. Hyper-Parameter Tuning

The learning process in deep learning models greatly relies on the values of hyperparameters that process the training procedure [38]. For the improved YOLOv6 model, we iteratively varied the hyperparameters to attain better accuracy and convergence performance. By adjusting various types of augmentations like mixup, flipping, scaling, and mosaic, the model was trained for 100 epochs. The base learning rate was lowered and momentum was adjusted to improve the accuracy consistently during training. Distribution Focal loss (DFL) is not adopted as the small and lightweight variant has no significant effect on the performance. A list of hyperparameters is shown in Table 1.

Table 1. A summary of hyperparameters used for training.

Training Parameter	Value
Optimizer	SGD
lr schedule	cosine
Use DFL	False
Batch size	32
Base learning rate lr_0	0.0036
Final learning rate lrf	0.13
Weight decay	0.00035
Momentum	0.849
Warmup epochs	2

2.4.4. SIOU Loss

The two main steps in an object detection model are to accurately predict the bounding box around ground truth and correctly classify the object's class in the bounding box. Bounding box regression loss defines the penalty that lies between the ground truth (B^{gt}) and prediction boxes (B). The loss is evaluated based on various parameters, i.e., the aspect ratio of boxes, distance between their centres and overlapping area. To evaluate the overlap of predicted and ground truth boxes, the intersection over union (IoU) metric is used. As evident, an object detector's mean accuracy precision (mAP) is based on IoU loss. The loss is constantly improved during training, leading to better detection and

classification. Several commonly used losses are CIoU, DIoU, and GIoU loss functions. However, in the proposed work, the efficiency of SIoU loss [39] function is utilized where the metrics are refined. The SIoU loss function after incorporating angle cost information to determine the mismatch direction between B and B^{gt} .

2.4.5. The Complete Model

Figure 4 gives the block diagram of our proposed model. Feature extraction is performed in the backbone using Repblock, RepConv, and cross-stage partial spatial pyramid pooling (CSPSPPF) block to enhance the network’s learning ability further. The vanishing gradient is addressed using the CSPSPPF block, which splits the feature part into two parts, and after applying the pooling operation, important spatial features are extracted. The split features are then merged via cross-stage hierarchy. The Repblock utilizes the GELU in place of ReLU as a non-linearity operator in the backbone block of the proposed model.

The YOLOv6s version of the network is used in this work, the relevant code of which can be found on GitHub. The model uses the EfficientRep structure as its backbone.

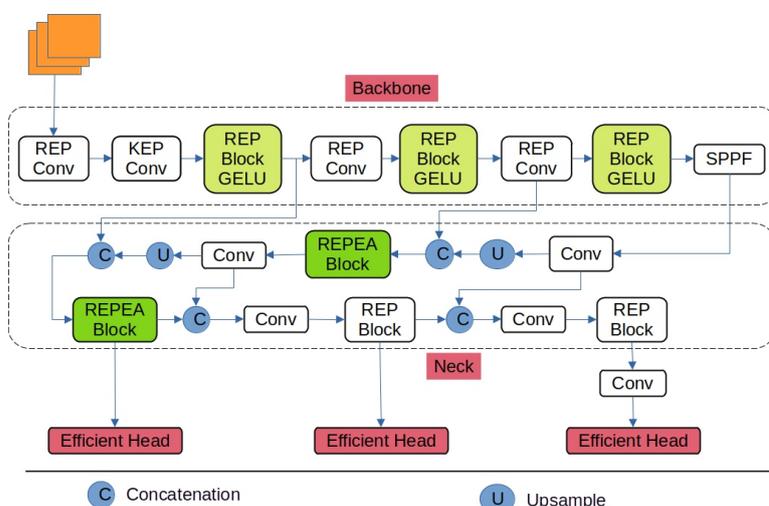


Figure 4. Proposed model for crop & fruit leaf disease detection.

The RepEA block replaces the RepBlock block in the neck to improve the detection of small targets under complex background conditions. The neck structure contains the CSP-styled stack of RepBlocks. The RepEA block is added after the Bifusion block, where the low-level, high-level, and current features are fused. Adding an attention mechanism will enhance the fused features aggregated by a path-aggregated network (PAN) at different scales.

3. The Self-Collected Crop & Fruit Disease Dataset

A challenging dataset is of prime importance for training a deep-learning model. Evaluating the detection model’s performance in challenging environmental conditions through smartphone usage could pave the way for the creation of a smartphone-assisted application intended for future use by farmers. For this purpose, a dataset comprising 1353 images and 3305 images after augmentation is collected. Various wheat, cotton, and mango diseases were captured from the southern Punjab region using a smartphone camera with resolution equal to 5 megapixels contained in Samsung SM-A217F smartphone). The dataset is composed of:

- 4 classes of wheat, namely yellow rust, brown rust, stem rust, smut & healthy wheat
- 3 classes of mango leaves namely anthracnose, nutrient deficient & healthy leaf) and
- 2 classes of cotton namely cotton leaf curl & healthy leaf.

Images were captured in various lighting situations, from various angles, and against different backgrounds. Additionally, the distance between the camera and the plants was

deliberately changed to introduce variation. The image was collected from various fields and orchids from March 2023 to August 2023. The diseased images were located, and after identifying the proper symptoms in the presence of a plant pathologist were captured. After careful filtering 87 blurred, improperly captured images were discarded. Some classes of wheat & cotton were further enriched by public datasets available on kaggle [40], CGIAR dataset [41] and CoSEV dataset(authors' public dataset) [9] respectively. A selection of images was also sourced from the internet to enhance the dataset's diversity. For this, we used the famous web sources of Google Images and Bing. Collecting images from various sources makes the dataset more challenging and will help train a more generalized deep-learning model. The detail of numbers of images captured by smartphone, sourced from the internet and public datasets in each class is mentioned in Table 2. Apart from 3 healthy classes, the dataset comprises wheat yellow rust, wheat brown rust, wheat stem rust, wheat smut, Mango anthracnose, mango nutrient deficiency, and cotton leaf curl.

Table 2. A summary of number of images of dataset.

Class	No. of Images by Smartphone	No. of Images from Internet Sourced	No of Images from Public Dataset
wheat yellow rust	128	58	35
stem rust	20	36	39
smut	194	51	-
mango Nutrient deficient	105	8	-
mango healthy	76	23	-
mango anthracnose	97	29	-
leaf rust	91	29	25
wheat healthy	82	19	27
cotton healthy	91	-	-
cotton curl	85	5	-

The snapshot of 10 classes of the collected dataset (marked as x -axis) is shown in Figure 5. Number of images (marked as y -axis) contained in each class is shown in Figure 6.

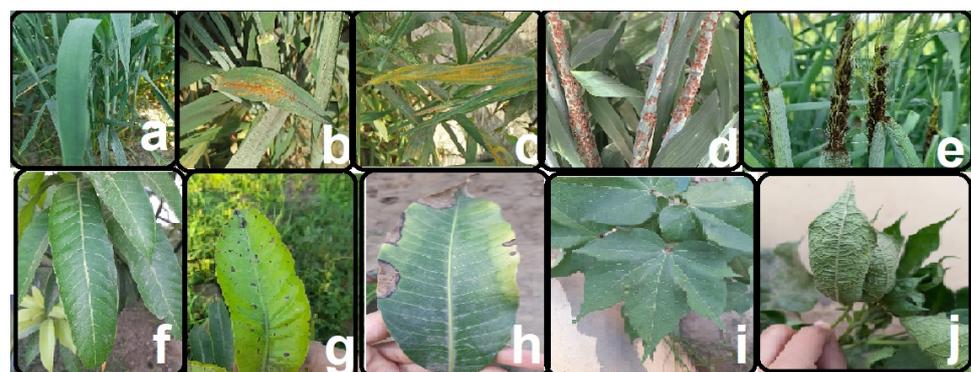


Figure 5. Self Collected dataset (a) Wheat healthy (b) Wheat brown rust (c) Wheat yellow rust (d) Wheat stem rust (e) Wheat smut (f) Mango healthy (g) Mango anthracnose (h) Mango nutrient deficient (i) Cotton healthy (j) Cotton Curl.

Diseased and healthy classes were manually annotated using the bounding box tool. Roboflow online tool is used to label the images using rectangular boxes. The labeling format for YOLOv6 is somewhat different from other versions of YOLO. The annotations of each image are saved in *.TXT* format. The annotation file contains information about the corresponding class and width, height, and coordinates of the bounding box. Labeled images were randomly split into train test and validation sets with a ratio of 88%: 5%:7% images. The directory structure containing train, test, and Validation images is linked in *.YAML* format.

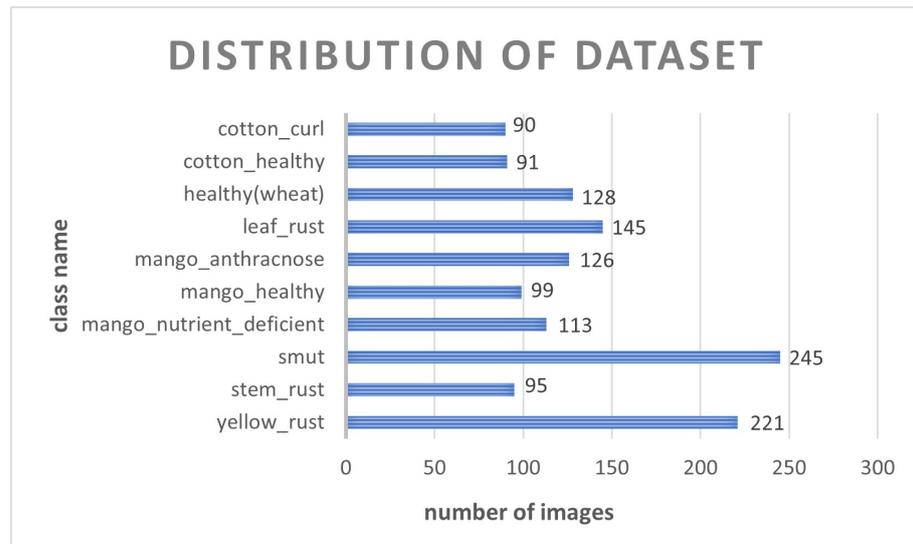


Figure 6. Visualizing the distribution of images in each class.

We have introduced several augmentations to enhance the effectiveness of the detection model and address the challenges posed by images taken in varying lighting, angles, and zoom conditions. These include vertical and horizontal image flipping, 25° image rotation, and a $\pm 25\%$ brightness adjustment. The effect of data enhancement process is shown in Figure 7. As a result of these augmentations, the total training images were increased three times.

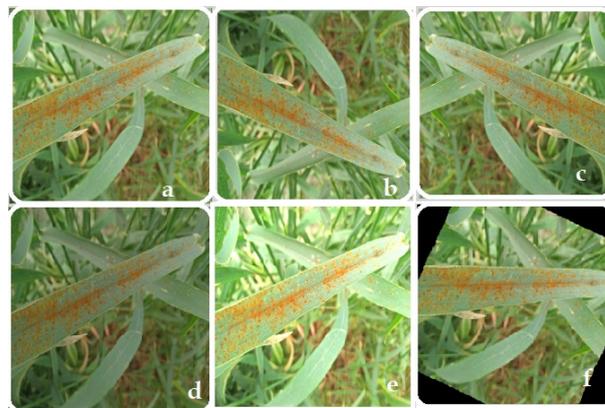


Figure 7. Data augmentation Steps (a) original image (b) Flip vertical (c) Flip horizontal (d) brightness -25% (e) brightness $+25\%$ (f) Rotate 25% .

4. Experimentation

The model is trained, validated, and tested using Colab Pro, Python 3.10 with GPU Nvidia T4 V100 to accelerate the training. Cuda 11.0 is used on the Pytorch deep learning framework. Several metrics are used to verify the model's effectiveness in detecting and classifying diseased symptoms. These evaluation metrics include precision (P), recall (R), mean accuracy precision (mAP), and detection time. Precision is the ratio of correct predictions and total predictions made by the model. The mean average precision can be expressed as shown in Equation (6).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

Higher mAP values indicate that the model makes more accurate predictions after training. Whereas, recall is the ratio of correct predictions to the number of ground truth detections as shown in Equation (7).

$$\text{Recall}(R) = \frac{\text{All_correct_detections}}{\text{Total_ground_truth_bounding_boxes}} \quad (7)$$

where AP_i is the average precision value of each class, mAP@ 50% is a commonly used metric in object detection models. The higher the value, the more accurate the detection. The proposed model is trained on the 2017 COCO dataset.

Initially, the mAP score was adjusted by fine-tuning of hyperparameter mentioned in Table 1. Moreover, GELU activation used in repblock effectively captures intricate feature details under varying light conditions. To improve the detection process further, the location of the ECA block was verified by adding various locations in the backbone and neck regions. Finally, best accuracy results with better recall are obtained after the addition of the REPEA block in the neck region of the model.

5. Results

Experiments are carried out to evaluate the performance of all other variants of YOLOv6 on the basis of accuracy, recall, training time, and computational cost. YOLOv6s version was chosen to be optimized for the specific problem addressed by this work due to its comparable performance at lower computational cost. As can be seen in Table 3, YOLOv6m took more time to train the the model so that it could perform detection accurately on our dataset with slightly improved performance metrics. The detection results of the proposed model are also compared with state-of-the-art object detection models, namely YOLOv5s, YOLOv7(base version), YOLOv8s, YOLO-NAS-s, YOLOS-s, and EfficinetDet.

All models are trained for 100 epochs using the default image size settings. A comparison of accuracy, parameters, and training times is given in Table 3. As can be seen, our proposed model outperforms all other techniques on the basis of the performance metrics considered for comparisons. YOLOv8s shows better recall performance but at the cost of increased training time and inferior mAP values at $IoU = 50\%$. The proposed model has 17.2 M parameters and 22 GFLOPS (Floating Point Operations), significantly lower than other models. Although the parameters of EfficinetDet are much lower at the cost of poor detection performance. To consider the requirements of real-time applications, the authors considered only small versions of all models. The model is trained for 100 epochs with 2.0 warmup epochs. With a batch size of 32, a cosine lr schedule is used having an initial learning rate of 0.0036. The declining loss curve for both classification and IoU loss is shown in Figure 8. This shows that as the training process continues, the targeted class is correctly localized and classified. The IoU loss gradually decreases as the training continues. However, the green line shows that classification loss sharply falls before the 20th epoch and between 90–100th epoch. Classification loss is the correctness of an object classified in the bounding box The improved performance of the proposed model can be seen in Figure 9, where the red line shows the average precision performance for all classes at IoU threshold 0.5 compared with the green line of mAP values for the proposed model. As can be seen, the detection ability greatly improves after the 20th epoch and finally tends to converge at better values in comparison with the baseline model.

Figure 10 shows the confusion matrix of the proposed model on the test dataset. As can be seen, the stem rust and wheat healthy classes are misclassified because of their higher similarity with the real-field background. Further, since the lesion area of stem rust is small, it worked as another reason for its misclassification. Few data images of the mango nutrient-deficient class were misclassified as mango anthracnose due to their similar symptoms. The wheat smut is the most missed class by the proposed detector probably because of its smaller sample size and smaller target area on the affected leaf. In some instances, brown rust is also misclassified as yellow rust. This is because under varying lighting conditions, symptoms of the two diseases become very similar. Mango healthy

and mango anthracnose are also misclassified from each other in the presence of cluttered backgrounds containing green leaves of the plant(s).

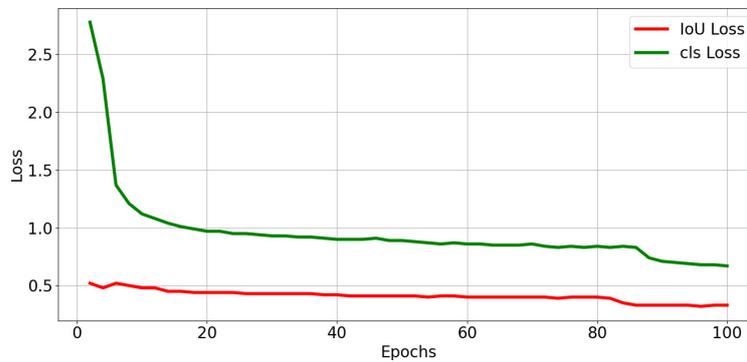


Figure 8. IoU and Classification loss curves during proposed model training.

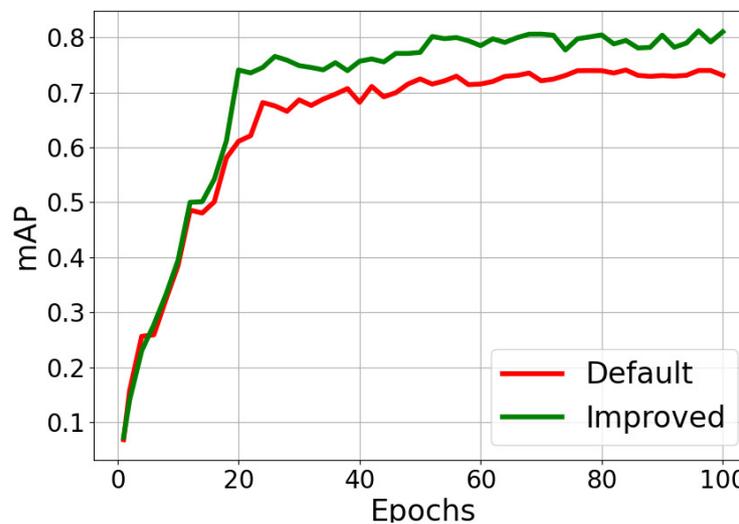


Figure 9. Comparison of mAP@50% for proposed model & default YOLOv6 model.

Table 3. Comparison of performance metric with different detection models.

Model	Image Size	Parameters (M)	Training Time (h)	Average Recall (AR)	mAP@50%
YOLOv5s	416	7.2	1.5	49.66	60.87
YOLOv7(base)	640	37	5.2	52.34	62.16
YOLOv8s	800	11.2	3.21	69.88	71.87
YOLOv8m	640	22.2	2.5	50.22	55.49
YOLOSs (DETR)	416	30.7	4.1	63.66	62.99
EfficientDet	512	3.9	4.55	44.98	50.48
YOLOv6m	640	21.2	2.37	65.59	75.29
YOLOv6s	640	17.2	1.48	67.41	73.14
Our Proposed Model	640	17.2	1.56	73.23	81.2

Figure 11 shows the precision-recall curves of all classes at 50% mAP. As the IoU threshold increases the detection accuracy of some classes decreases. The curves of stem rust and healthy(wheat) are more affected due to similarities of their symptoms with other classes. Cotton curl and cotton healthy obtain higher scores in terms of precision and recall as most of their images have high contrast image conditions. The precision and average recall of each class is also shown in Table 4. Almost all the classes are detected with

reasonably large values of mAP. Stem rust and wheat healthy classes are either not correctly detected or missed by the model. This is due to the similarity of diseased symptoms with the background.

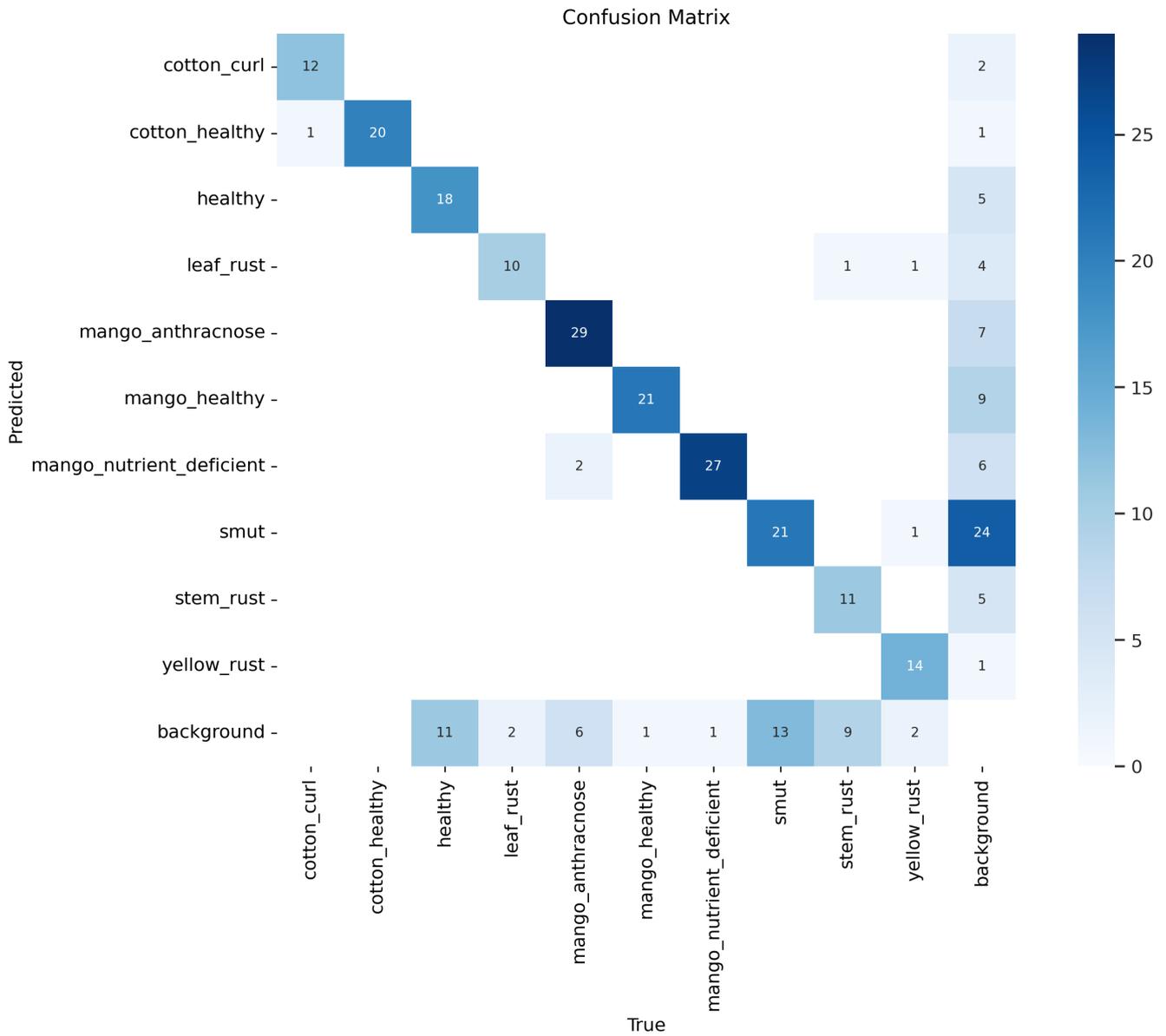


Figure 10. Confusion matrix of the proposed model on test dataset.

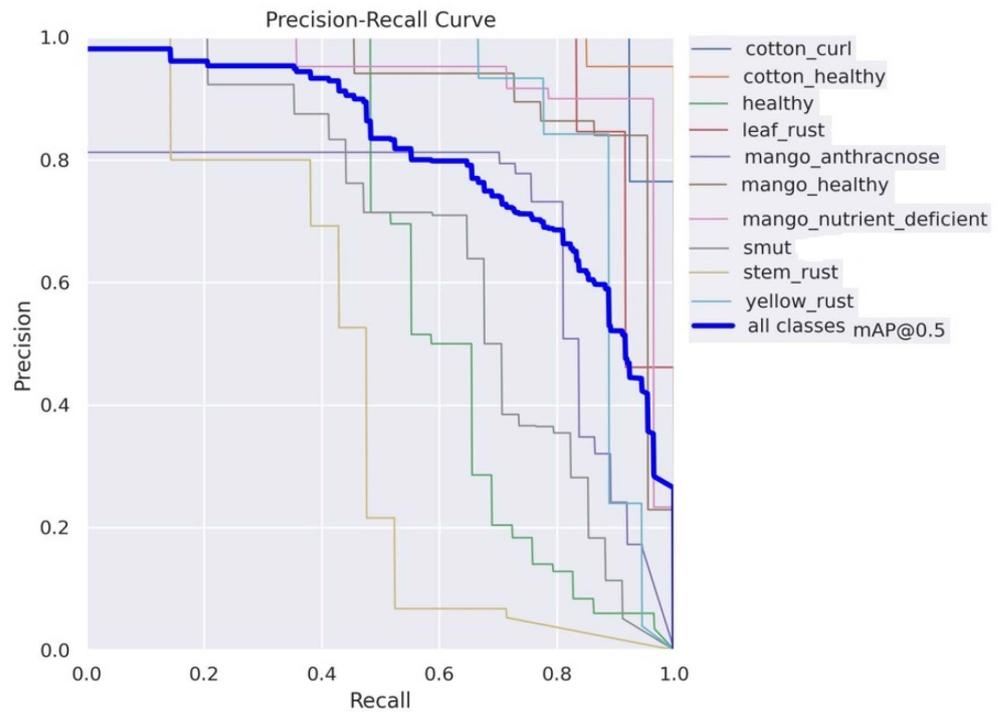


Figure 11. Precision-Recall curve of all 10 classes @ IoU threshold of 0.5.

Table 4. A comprehensive summary of validation results of the proposed model.

Class	Images	Instances	AP	AR	mAP @ 0.50	mAP @ 0.50:0.95
All	152	234	0.849	0.732	0.8124	0.472
cotton_curl	152	13	0.788	0.923	0.979	0.735
cotton_healthy	152	20	0.922	1	0.988	0.68
healthy	152	29	0.774	0.483	0.626	0.31
leaf_rust	152	12	0.959	0.833	0.94	0.501
mango_anthracnose	152	37	0.792	0.721	0.701	0.413
mango_healthy	152	22	0.841	0.864	0.916	0.655
mango_nutrient_deficient	152	28	0.876	0.964	0.931	0.61
smut	152	34	0.767	0.482	0.745	0.202
stem_rust	152	21	0.753	0.381	0.422	0.223
yellow_rust	152	18	0.987	0.667	0.876	0.39

5.1. Ablation Experiments

In an attempt to verify the effectiveness of the proposed model, numerous ablation studies have been conducted. The experimental settings and dataset version are kept the same while performing the ablation experiments to maintain comparability. The use of ✓ in Table 5 refers to the use of that method.

Table 5. Results of an ablation study conducted during training the model with varying schemes.

YOLOv6s	Fine-Tuning	SIOU	GELU	REPEA	mAP(%)
✓	×	×	×	×	73.14
✓	✓	×	×	×	75.48
✓	✓	✓	×	×	76.97
✓	✓	✓	✓	×	79.03
✓	✓	✓	✓	✓	81.2

As can be seen in Table 5, the training accuracy is improved after fine-tuning the baseline model. An optimization performance is achieved by changing the bounding box

regression loss from *GIoU* loss to *SIoU* loss. A 2.06% increase in mAP is obtained by replacing the non-linearity function in the model backbone with GELU as compared to ReLU activation. Better convergence is obtained as GELU is non-convex, non-monotonic, and is not linear in the positive domain, in contrast to the ReLU activation function. A further increase in accuracy is observed after integrating efficient channel attention in the neck of the YOLOv6s model.

Consequently, compared to the baseline model, the improved model showed an overall increase of 7.92% in the mAP@50% score.

5.2. Discussions on Results

Some sample instances where the proposed deep learning model detects the leaf diseases are given in Figure 12.

In Figure 12b, it can be seen that the targeted class is not only detected with a better confidence score but also some targets that were treated as background in Figure 12a are also detected. In Figure 12d target is detected with better confidence and, in turn, with improved mAP performance. In Figure 12f, target localization is further improved due to the use of SIOU regression loss.

In Figure 12g, the diseased area is not detected as it is similar to the background and lightning conditions make the situation more difficult. But as can be seen in Figure 12h, the leaf rust is detected with an improved confidence score; the leaf rust class present in the background is also detected by the proposed model. As shown in Figures 10 and 12, there are several missed and false detections, which happen mostly in images with similar or cluttered background conditions and low-resolution images with blurry ground truth. However, the model detection performance was found to be superior when compared with other models. The accuracy and recall of YOLOv8s were found closer to comparable training time but still, the performance of the proposed model was above par in comparison. It can be concluded hence, that when dealing with real-field images captured via smartphone, almost all models exhibit poor performance, which is why it is still a challenge for researchers.

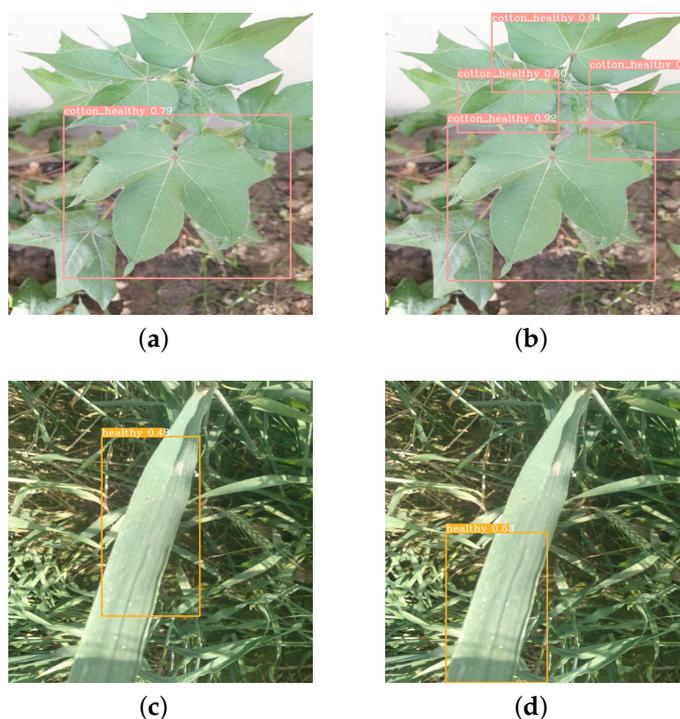


Figure 12. Cont.

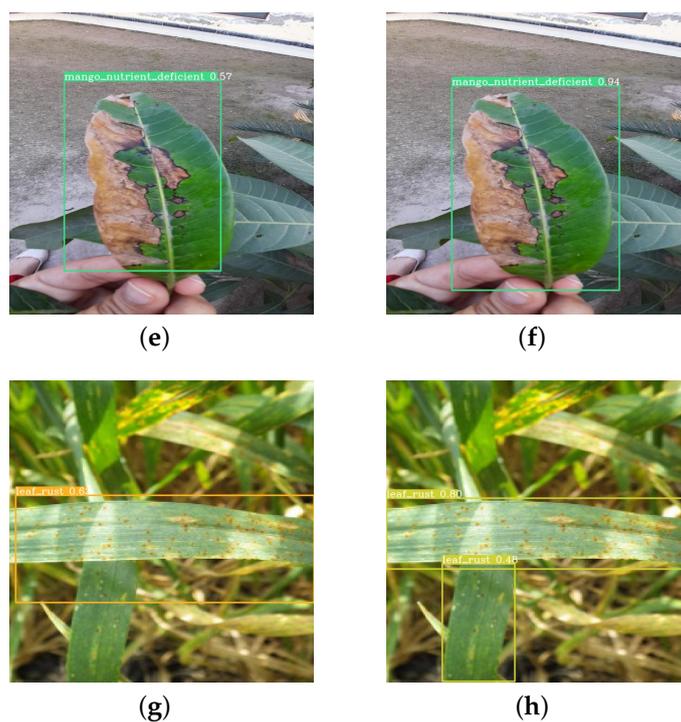


Figure 12. Detection results on test dataset. (a,c,e,g) results of default YOLOV6 model. (b,d,f,h) Results of Improved YOLOV6 model.

6. Conclusions & Future Work

The study aims to introduce an enhanced approach for identifying diseased areas on plants. Numerous studies in the field utilize various computer vision techniques to classify and locate plant diseases on public datasets. However, advanced recognition models often struggle to detect symptoms in intricate field environments. Challenges arise from variable lighting conditions, complex and similar backgrounds, variable lesion/diseased areas, and low contrast, making the detection process particularly challenging. In this regard, we have proposed an improved model utilizing the Efficient channel attention mechanism integrated into the baseline model of YOLOv6. The regression and localization task is further improved via fine-tuning and the use of the SIOU loss function. To improve the detection performance further GELU function is incorporated as a non-linearity function. The mAP score of the proposed model is 81.2% and an Average recall of 73.2% after 1.56 h of training. The requirement of a robust real-time detection model is better accuracy in a shorter time and lesser computational cost. The results obtained using the proposed model are also compared with other recent small and/or baseline versions of various models and are found superior in terms of recall, accuracy, and training time in complex environmental conditions. As the proposed dataset comprises images with varying resolution, the robustness of the model in the detection of small lesion areas is better compared to other models.

However, the detection accuracy suffered due to the imbalance of the dataset. Several classes are left undetected due to the low contrast of images. Wheat smut images are low in number, so the imbalance resulted in low precision. Moreover, due to varying lighting conditions and disease severity, few classes of yellow rust are falsely detected as brown leaf rust.

In the future, we intend to enhance our dataset in terms of the number of images and classes to make it further richer and closer to the real-field conditions. To make the model widely applicable to different types of plant disease detection areas, we wish to extend the number of images by covering a wide variety of crops. In addition, further studies will focus on gathering the environmental information to construct a multisource

fusion model to gather information about humidity, temperature, and soil information to predict favorable conditions for a particular pathogen. That will make the early diagnosis of infected crops easier.

Author Contributions: Conceptualization, S.K.N. and M.A.; methodology, S.K.N. and M.A.Q.; software, S.K.N. and A.M.; validation, S.K.N., A.M., M.A. and M.A.Q.; formal analysis, S.K.N.; resources, T.A.; data curation, S.K.N. and M.A.; writing—original draft preparation, S.K.N., M.A.Q. and A.M.; writing—review and editing, M.A. and T.A.; supervision, M.A. and M.A.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors appreciate the support and help extended by the Department of Plant Pathology, The Islamia University of Bahawalpur in identifying and annotating the images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Usman, M. Contribution of agriculture sector in the GDP growth rate of Pakistan. *J. Glob. Econ.* **2016**, *4*, 1–3.
2. Shah, H.; Siderius, C.; Hellegers, P. Cost and effectiveness of in-season strategies for coping with weather variability in Pakistan's agriculture. *Agric. Syst.* **2020**, *178*, 102746. [[CrossRef](#)]
3. Akhtar, C. Principal diseases of major crops in Pakistan with reference to genetic resistance. In *Genetic Diversity in Plants*; Springer: Berlin/Heidelberg, Germany, 1977; pp. 179–191.
4. Cheema, M.J.M.; Iqbal, T.; Daccache, A.; Hussain, S.; Awais, M. Precision agriculture technologies: Present adoption and future strategies. In *Precision Agriculture*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 231–250.
5. Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A. Handling similar looking disease symptoms in plants using dilation and feature reuse. *J. Intell. Fuzzy Syst.* **2023**, *45*, 1–16. [[CrossRef](#)]
6. Lu, J.; Tan, L.; Jiang, H. Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture* **2021**, *11*, 707. [[CrossRef](#)]
7. Saleem, R.; Shah, J.H.; Sharif, M.; Ansari, G.J. Mango Leaf Disease Identification Using Fully Resolution Convolutional Network. *Comput. Mater. Contin.* **2021**, *69*, 3581–3601. [[CrossRef](#)]
8. Sangeetha, R.; Logeshwaran, J.; Rocher, J.; Lloret, J. An Improved Agro Deep Learning Model for Detection of Panama Wilts Disease in Banana Leaves. *AgriEngineering* **2023**, *5*, 660–679. [[CrossRef](#)]
9. Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A. Handling severity levels of multiple co-occurring cotton plant diseases using improved YOLOX model. *IEEE Access* **2022**, *10*, 134811–134825. [[CrossRef](#)]
10. Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A. Use of deep learning techniques for identification of plant leaf stresses: A review. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100443. [[CrossRef](#)]
11. Ngongoma, M.S.; Kabeya, M.; Moloi, K. A Review of Plant Disease Detection Systems for Farming Applications. *Appl. Sci.* **2023**, *13*, 5982. [[CrossRef](#)]
12. Du, X.; Cheng, H.; Ma, Z.; Lu, W.; Wang, M.; Meng, Z.; Jiang, C.; Hong, F. DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Comput. Electron. Agric.* **2023**, *214*, 108304. [[CrossRef](#)]
13. Li, W.; Zhang, L.; Wu, C.; Cui, Z.; Niu, C. A new lightweight deep neural network for surface scratch detection. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 1999–2015. [[CrossRef](#)] [[PubMed](#)]
14. Maheswaran, S.; Indhumathi, N.; Dhanalakshmi, S.; Nandita, S.; Mohammed Shafiq, I.; Rithka, P. Identification and Classification of Groundnut Leaf Disease Using Convolutional Neural Network. In Proceedings of the International Conference on Computational Intelligence in Data Science, Koceli, Turkey, 16–17 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 251–270.
15. Khirade, S.D.; Patil, A. Plant disease detection using image processing. In Proceedings of the 2015 International Conference on Computing Communication Control and Automation, IEEE, Pune, India, 26–27 February 2015; pp. 768–771.
16. Paymode, A.S.; Malode, V.B. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artif. Intell. Agric.* **2022**, *6*, 23–33. [[CrossRef](#)]
17. Saleem, M.H.; Khanchi, S.; Potgieter, J.; Arif, K.M. Image-based plant disease identification by deep learning meta-architectures. *Plants* **2020**, *9*, 1451. [[CrossRef](#)] [[PubMed](#)]
18. Wang, J.; Yu, L.; Yang, J.; Dong, H. Dba_ssd: A novel end-to-end object detection algorithm applied to plant disease detection. *Information* **2021**, *12*, 474. [[CrossRef](#)]
19. Alqahtani, Y.; Nawaz, M.; Nazir, T.; Javed, A.; Jeribi, F.; Tahir, A. An improved deep learning approach for localization and Recognition of plant leaf diseases. *Expert Syst. Appl.* **2023**, *230*, 120717. [[CrossRef](#)]

20. Chowdhury, M.E.; Rahman, T.; Khandakar, A.; Ayari, M.A.; Khan, A.U.; Khan, M.S.; Al-Emadi, N.; Reaz, M.B.I.; Islam, M.T.; Ali, S.H.M. Automatic and reliable leaf disease detection using deep learning techniques. *AgriEngineering* **2021**, *3*, 294–312. [[CrossRef](#)]
21. Wang, H.; Shang, S.; Wang, D.; He, X.; Feng, K.; Zhu, H. Plant disease detection and classification method based on the optimized lightweight YOLOv5 model. *Agriculture* **2022**, *12*, 931. [[CrossRef](#)]
22. Zhang, Y.; Zhou, G.; Chen, A.; He, M.; Li, J.; Hu, Y. A precise apple leaf diseases detection using BCTNet under unconstrained environments. *Comput. Electron. Agric.* **2023**, *212*, 108132. [[CrossRef](#)]
23. Zhao, W.; Wu, D.; Zheng, X. Detection of Chrysanthemums Inflorescence Based on Improved CR-YOLOv5s Algorithm. *Sensors* **2023**, *23*, 4234. [[CrossRef](#)]
24. Piao, Z.; Wang, J.; Tang, L.; Zhao, B.; Wang, W. AccLoc: Anchor-Free and two-stage detector for accurate object localization. *Pattern Recognit.* **2022**, *126*, 108523. [[CrossRef](#)]
25. Kaur, P.; Harnal, S.; Gautam, V.; Singh, M.P.; Singh, S.P. An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105210. [[CrossRef](#)]
26. Liu, J.; Wang, X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front. Plant Sci.* **2020**, *11*, 898. [[CrossRef](#)] [[PubMed](#)]
27. Soeb, M.J.A.; Jubayer, M.F.; Tarin, T.A.; Al Mamun, M.R.; Ruhad, F.M.; Parven, A.; Mubarak, N.M.; Karri, S.L.; Meftaul, I.M. Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci. Rep.* **2023**, *13*, 6078. [[CrossRef](#)] [[PubMed](#)]
28. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
29. Norkobil Saydirasulovich, S.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.I. A YOLOv6-based improved fire detection approach for smart city environments. *Sensors* **2023**, *23*, 3161. [[CrossRef](#)] [[PubMed](#)]
30. Zhang, H.; Wang, Y.; Dayoub, F.; Sunderhauf, N. Varifocalnet: An iou-aware dense object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 8514–8523.
31. Weng, K.; Chu, X.; Xu, X.; Huang, J.; Wei, X. EfficientRep: An Efficient Repvgg-style ConvNets with Hardware-aware Neural Network Design. *arXiv* **2023**, arXiv:2302.00386.
32. Zhang, J.L.; Su, W.H.; Zhang, H.Y.; Peng, Y. SE-YOLOv5x: An optimized model based on transfer learning and visual attention mechanism for identifying and localizing weeds and vegetables. *Agronomy* **2022**, *12*, 2061. [[CrossRef](#)]
33. Zhao, Y.; Chen, J.; Xu, X.; Lei, J.; Zhou, W. SEV-Net: Residual network embedded with attention mechanism for plant disease severity detection. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6161. [[CrossRef](#)]
34. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
35. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13733–13742.
36. Lee, M. Mathematical analysis and performance evaluation of the gelu activation function in deep learning. *J. Math.* **2023**, *2023*, 4229924. [[CrossRef](#)]
37. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
38. Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A. Overfitting mitigation analysis in deep learning models for plant leaf disease recognition. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), IEEE, Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–5.
39. Gevorgyan, Z. SIOU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
40. GETCH, O. Wheat Leaf Dataset. 2021. Available online: <https://www.kaggle.com/datasets/olyadgetch/wheat-leaf-dataset> (accessed on 3 November 2023).
41. Alharbi, A.; Khan, M.U.G.; Tayyaba, B. Wheat Disease Classification using Continual Learning. *IEEE Access* **2023**, *11*, 90016–90026. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.