Scientific Reports
https://doi.org/10.1038/s41598-026-45575-1

Article in Press

A novel approach for disease and pests detection in potato production system based on deep learning

Received: 18 November 2025

Accepted: 19 March 2026

Published online: 02 April 2026

Cite this article as: Abbas A., Rehman S.U., Mahmood K. *et al.* A novel approach for disease and pests detection in potato production system based on deep learning. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-45575-1

Ahmed Abbas, Saif Ur Rehman, Khalid Mahmood, Santos Gracia Villar, Luis Alonso Dzul Lopez, Aseel Smerat & Imran Ashraf

A Novel Approach for Disease and Pests Detection in Potato Production System based on Deep Learning

Ahmed Abbas[1], Saif Ur Rehman[1*], Khalid Mahmood[2], Santos Gracia Villar[3,4,5,6], Luis Alonso Dzul Lopez[3,4,7,8], Aseel Smerat[9,10], Imran Ashraf[11*]

[1]PMAS-Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan.
[2]Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan.
[3]Universidad Europea del Atlntico., Isabel Torres 21, Santander, 39011, Spain.
[4]Universidad Internacional Iberoamericana, Campeche, 24560, Mexico.
[5]Universidad Internacional Iberoamericana Arecibo, Puerto Rico, 00613, USA.
[6]Universidade Internacional do Cuanza, Cuito, Bie,, Angola.
[7]Fundacion Universitaria Internacional de Colombia, Bogota, Colombia.
[8]Universidad de La Romana, La Romana, Republica Dominicana.
[9]Faculty of Educational Sciences, Al-Ahliyya Amman University, Amman, 19328, Jordan.
[10]Department of Biosciences, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India.
[11]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea.

*Corresponding author(s). E-mail(s): saif@uaar.edu.pk; imranashraf92@gmail.com;
Contributing authors: ahmeduiit@yahoo.com; khalid@gu.edu.pk; santos.gracia@uneatlantico.es; luis.dzul@unini.edu.mx; smerat.2020@gmail.com;

Abstract

Vulnerability of potato crops to diseases and pest infestation can affect its quality and lead to significant yield losses. Timely detection of such diseases can help take effective decisions. For this purpose, a deep learning-based object detection framework is designed in this study to identify and classify major potato diseases and pests under real-world field conditions. A total of 2,688 field images were collected from two research farms in Punjab, Pakistan, across multiple growth stages in various seasonal conditions. Excluding 285 symptoms-free images from the earliest collection led to 2,403 images which were annotated into four biotic-stress classes: blight disease (n=630), leaf spot disease (n=370), leafroll virus (viral symptom complex; n=888), and Colorado potato beetle (larvae/adults; n=515), indicating class imbalance. Several state-of-the-art models were used including YOLOv8 variants (n/s/m), YOLOv7, YOLOv5, and Faster R-CNN, and the results are discussed in relation to recent potato disease classification studies involving cropped leaf images. Stratified splitting (70% training, 20% validation, 10% testing) was applied to preserve class distribution across all subsets. YOLOv8-medium achieve the best performance with mean average precision (mAP)@0.5 of 98% on the held-out test images. Results for stable 5-fold cross-validation show a mean mAP@0.5 of 97.8%, which offers a balance between accuracy and inference time. Model robustness was evaluated using 5-fold cross-validation and repeated training with different random seeds, showing a low variance of ±0.4% mAP. Results demonstrate promising outcomes under the real-world field conditions, while, broader cross-region and cross-season validation is intended for the future.

Keywords: Pests detection; disease detection; convolutional neural network; object detection; object classification; deep learning

1 Introduction

Agriculture is a sector of great importance in the economies of countries like Pakistan, as the economy is directly or indirectly dependent on agriculture [1, 2]. It has been regarded as the key source of revenue for Pakistan's huge population. Over the years, for the farmers and consumers alike, potato has become an important crop in Pakistan. Potato plant production is the fourth most significant crop in terms of production volume, thus providing higher yields and favorable profits to farmers. Potato has more protein, iron, and nutrients than other vegetables in the usual diet. Potato is also high in nutrients such as thiamine, fiber, and niacin. So, getting high yield and good quality corp is very important.

Potato crop is affected by different kind of leaf diseases and pests; thus, early disease and pest detection are critical in the agriculture sector [3]. When potato crop is attacked by disease and pests, it is critical to take appropriate measures to manage crop plant quality and yield. Despite how simple it might appear, identifying potato plant disease is a difficult process. To detect potato plant leaf disease and pests, skilled staff expertise are required, which is not an easy task. Although the diseases appear similar in their early stages, they are difficult to differentiate with the naked eye and take time to diagnose. As a result, farmers would need to contact experienced specialists for plant identification, which is costly, time-consuming, and occasionally erroneous. Deep learning (DL) has become a key research area in this regard, and it is used extensively in agriculture for reasons, ranging from fruitlet categorization to weed identification [4]. DL approaches are used in agricultural research due to their capacity to automatically learn attributes from image datasets, as well as their greater levels of accuracy and speed than prior techniques [5–7].

In recent years, machine learning (ML), especially DL models, enabled more reliable agricultural decision support systems. DL models can learn complex visual patterns from large datasets under variable field conditions, and support scalable early-warning and intervention systems [8, 9]. The convolutional neural network (CNN) is the most widely used DL architecture. It was created especially for object identification, disease diagnosis, medical image analysis, image recognition in videos [6], the analysis of soil [2], the detection of pests, etc.

Despite several works on potato crop disease detection, existing approaches often rely on controlled or leaf-level classification, and can struggle in real field scenes with background clutter, illumination variability, leaf overlap/occlusion, and small pest instances. By jointly detecting both diseases and pests with localization in a single pass, the framework targets these practical gaps for in-field monitoring. Compared to DL models like CNN, sophisticated you look only once (YOLO) model has shown better performance, robustness and suitability for real-time applications. The essence of the YOLO detection algorithm revolves around its compact size and efficient computational speed. YOLO has been widely employed to recognize, classify, segment, and track objects in real time. The swiftness of YOLO is attributed to its ability to process an entire image in one go, swiftly providing the final detection output. This characteristic also enables YOLO to perform real-time detection. An advantage of YOLO is its utilization of the entire image for detection, enabling the encoding of global information and reducing the likelihood of mistakenly identifying the background as an object during the detection process [10].

In this study, different versions of the YOLO model are used to identify diseases and pests on potato plant leaves using the collected images. The success of the detection method depends on how accurate and fast it is. The current approach helps farmers quickly and correctly spot problems on leaves and warns them about diseases before they spread. The problem is formulated as object detection: the model predicts bounding boxes and class labels for visible disease lesions and pest bodies in field imagery, rather than only classifying cropped leaves. For real-field applicability, portable video cameras were used for data collection considering the growth stage of the potato plant. Field workers with expert knowledge are essential for effectively operating such a system. In this study, on site discussions were conducted with experienced plant specialists. The experts identified different indicators during early plant examinations that helped detect unhealthy or irregular plants in the early stages of development. As the plants matured into the mid growth stage, factors such as overlapping leaves from neighboring plants and the difficulty in detecting plants at early developmental stages became more apparent.

The proposed approach is divided into four stages. In the first stage, the data is gathered from various plant leaves that are affected by various diseases. All these images are captured as color images. The noise from the photographs is reduced in the first stage. In the second stage, the images are preprocessed, and augmentation is applied. In the third stage, annotation of the images is done using the LabelImg tool. Finally, several YOLO variants such as YOLOv8 medium (YOLOv8-m), YOLOv8 small (YOLOv8-s), YOLOv8 nano (YOLOv8-n), YOLOv7 small (YOLOv7-s), and Faster R-CNN are trained and compared based on accuracy, other metrics. In brief, this study offers several key contributions to the field of precision agriculture and plant disease detection.

- **Development of a Custom, Field-collected Dataset:** Unlike many studies that rely on publicly available or lab-controlled datasets, this study gathered a real-world dataset from two distinct research farms under varied light, environmental, and plant growth conditions.
- **Comprehensive Benchmarking of Modern YOLO Variants:** Owing to the importance of accurate leaf disease and pest detection in potatoes, this study provides a detailed comparative analysis of multiple YOLOv8 variants, YOLOv7, YOLOv5, and Faster-RCNN.
- **Performance Analysis of YOLOv8-m:** The study highlights and explains the architectural reasons behind YOLOv8-m's superior performance (98% mAP), including its anchor-free design and improved feature extraction capability over its lighter counterparts.
- **Detection Framework for Real-time Deployment:** The model is designed to work in real-time with low inference time, demonstrating practical applicability in the field using portable camera systems.
- **Discussion of Future Research Work:** The study critically acknowledges the geographic constraints of its dataset and outlines the necessity of cross-regional validation, providing a more grounded perspective on real-world deployment.

In contrast to several prior studies that either focus solely on diseases or pests, or that operate on controlled or lab-acquired datasets, this work addresses biotic stresses jointly and under realistic field variability. Recent advances such as hybrid EfficientNet-ViT models, self-supervised learning (SSL) pipelines, multimodal frameworks like PotatoGPT, and explainable CNNs have demonstrated strong performance for potato disease classification at the leaf level, but they do not directly tackle joint disease-pest detection in cluttered field imagery or real-time object detection. This work addresses multi-object detection under real-world field conditions with background clutter, occlusion, and illumination variability. Because of differences in datasets and problem formulation, direct numerical comparison must be interpreted cautiously. By positioning YOLO-based detectors within this broader context, this study provides a practical and comparative baseline for future methods, including Transformer-based detectors (e.g., DETR/ViTDet), SSL-pretrained backbones, and ensemble or stacking schemes that integrate CNN and ViT features for enhanced robustness.

2 Literature Review

Recent advancements in DL have led to the exploration of ensemble and stacking methodologies to improve the accuracy and generalizability of plant disease detection models. For instance, [11] introduced a hybrid model combining EfficientNetV2B3 and vision transformer (ViT), which resulted in a test accuracy of 85.06% on a challenging multi-class classification dataset of potato leaf images. While the model was highly effective in identifying bacterial and fungal infections, it showed limited accuracy of around 60.7% in detecting pest-related symptoms, indicating the persistent challenge of distinguishing between disease and pest damage due to overlapping visual patterns.

In [12], the authors combined deep CNN-based feature extraction with feature concatenation, wrapper-based feature selection, and support vector machines (SVM) classification, achieving nearly 99% accuracy on early and late blight images. The hybrid classical-deep approach demonstrated that effective feature engineering paired with ML classifiers could offer high accuracy and generalizability. Additionally, the authors used transfer learning techniques where a fine-tuned DenseNet201 model achieved 96% validation accuracy and 92.5% accuracy on test data, even with a limited dataset. This supports the efficiency of pre-trained architectures for agricultural disease detection [13]. Other end-to-end CNN approaches have also proven robust; the study [14] reported 99.1% accuracy in classifying healthy leaves alongside early and late blight symptoms.

Further comparative evaluations of several DL models such as VGG16, VGG19, MobileNetV2, ResNet50, and AlexNet showed that ResNet50 provided the highest accuracy of around 97% and specificity of 98%, outperforming VGG-based architectures, which struggled with misclassifying healthy samples [12]. However, pest identification still proves more difficult than disease detection. To address this problem, a pest-specific DL framework, PotatoPestNet, was introduced. Built on a CT-InceptionV3-RS backbone, the model targeted eight common pest categories and achieved over 91% across all performance metrics, demonstrating the advantage of specialized pest detection pipelines [15]. In [16], the authors integrated explainable artificial intelligence (XAI) techniques, specifically saliency maps with a ResNet-9 architecture. In addition to achieving a 99.25% classification accuracy, the proposed approach also enhanced transparency by highlighting the regions of input images that contribute more to the model's decisions.

Another recent research work by Abbas et al. [17] introduced an AI framework for disease detection in potato plants based on a CNN model. Using over 20,000 PlantVillage images, the study classified potato leaf diseases across 15 classes, including healthy, fungal, and bacterial diseases, with 98.29% training accuracy and 98.03% test accuracy, demonstrating robustness on a large, diverse dataset relevant to Pakistan agricultural conditions. The study [18] explored real-time detection of the Colorado potato beetle using drone-captured field images. The dataset comprised 9,000 images (adults, larvae, no beetles), and six DL models (MobileNet, InceptionV3, ResNet101, AlexNet, DenseNet121, and Xception) were evaluated. While average accuracy was 96.3%, performance dropped to 92.95% when tested on independent field data, highlighting challenges in domain generalization for pest detection.

The study [19] proposed a multimodal AI model (MSC ResViT + MSC TextCNN + CT CNN) which is combined with a large language model interface called PotatoGPT. The system achieved 98.43% accuracy on disease classification (early vs late blight), surpassing image-only models. It also generalized to other Solanaceae crops and delivered integrated disease diagnosis and management advice via text-based interaction, representing a significant leap toward intelligent, user-friendly field tools. Poyato et al. [20] tackled the small-data challenge in field-based potato disease detection. Using high-resolution RGB field images and a patch-based CNN approach with focal loss and tailored data augmentation, the model reliably detected early symptoms of late blight even under low-data conditions, correctly identifying all blight cases in the test set. This demonstrates strong potential for real-world, resource-constrained agricultural deployments.

ViT-based approaches have reported better results for plant disease detection. For example, [21] used hybrid approach comprising ViT with Efficient- NetVnB3 to get improved detection accuracy for potato disease detection. Similarly, the authors in [22] used ViT for detection and classification of diseases from potato leaves. The study [23] used ViT with synthetic data from stable diffusion and reported better results for detecting various diseases.

Mamba-based models also made their way for plant disease detection. The authors [24] introduced VMamba which combined selective scanning and diffusion methods. It showed better performance, and is particularly suited for smaller datasets. Another research [25] developed Mamba-YOLO-ML model by integrating Mamba with Haar wavelet downsampling to get better results for leaf diseases.

Besides supervised ML, semi-supervised learning (SSL)-based approaches have also been utilized for potato disease and pest detection. For example, [26] used consistency regularized-based SSL and employed the mean teacher framework on unlabeled dataset. Similarly, [27] combined contrastive learning with few-shot strategy in the context of plant disease detection. Encouraging results have been reported using only 30 training images. Generative adversarial network (GAN)-assisted data augmentation has also been investigate for potato disease detection [28]. Recently, it has been reported that using SSL and clustering to combine labeled and unlabeled data might lead to enhanced performance [29].

The discussed studies underscore the potential of ensemble and stacking methods to enhance model performance, particularly in scenarios with diverse environmental conditions and limited labeled data. Incorporating such methodologies could address the challenges of generalizability and robustness in plant disease detection systems. Analysis of existing literature reveals that prior studies have used CNNs, transfer learning, e.g., DenseNet201, ResNet50, etc., and hybrid models for disease detection with varying accuracy. However, pest detection remains a significant challenge due to overlapping visual symptoms, varying light conditions and change in symptoms across various growth states. In addition, generalizability across various climatic zones is limited. The challenge of distinguishing biotic vs abiotic stresses also requires further investigation. Thus, while high accuracy has been reported in controlled/lab settings, challenges remain in scalability, robustness to new regions, pest-specific detection, and early/late blight differentiation. In view of these shortcoming, this study addresses a few of these issues by

- Creating a custom, field-collected dataset under varied natural conditions, unlike many prior lab-based datasets.
- Conducting comprehensive benchmarking of multiple YOLO versions and Faster R-CNN to assess robustness.
- Evaluating YOLOv8-m which balances speed and accuracy, outperforming prior models.
- Applying data augmentation, normalization, and expert annotation to enhance dataset quality.
- Explicitly discussing limitations like regional bias and proposing cross-regional validation for scalability.

3 Materials and Methods

The objective of this study is to test YOLO's ability to detect disease and pests from image of real potato crop. Figure 1 depicts the workflow of the suggested approach.

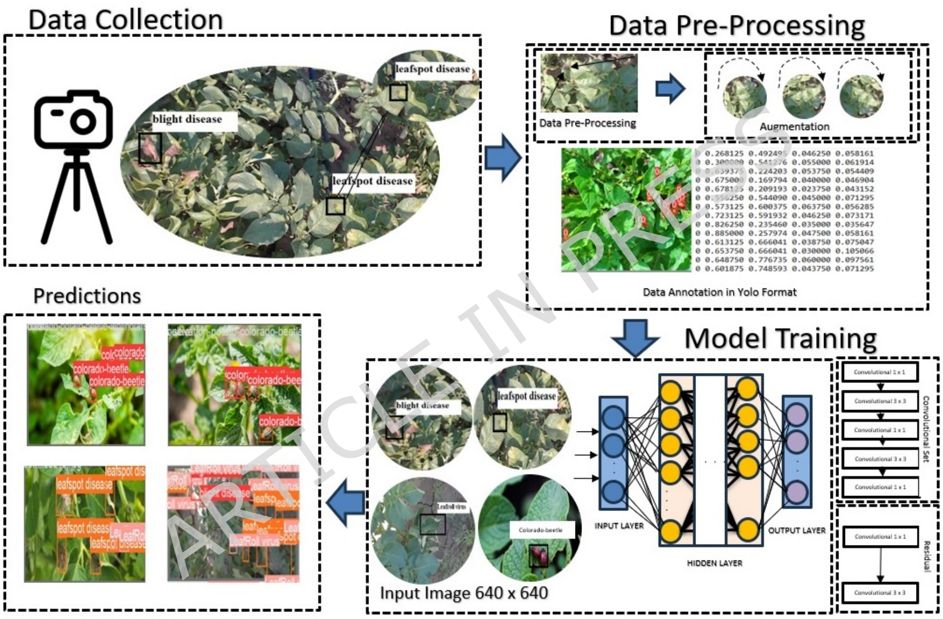


Fig. 1: Suggested disease and pest detection system's process.

High-quality and larger dataset is essential for constructing efficient DL models, as the models' performance heavily relies on the dataset's quality, quantity, and relevance. As a result, capturing high-quality images is the first and most important stage in this study. Figure 1 depicts the complete procedure, including image capture, model training, and results analysis. Additionally, various additional processes were carried out to ensure good model fit and improved accuracy. These processes are briefly described here.

i. Images for potato plant disease and pests were methodically collected from a variety of sources, including the University Research Farm Koont, the Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, and the University of Faisalabad, Okara campus. The crop images were collected at various phases of growth (initial, medium, and mature), during different days, under different climate conditions.
ii. After the data collection phase, image preprocessing was carried out, which encompassed resizing the collected images to ensure uniformity and consistency.
iii. Data augmentation techniques, including flipping, rotating, scaling, and cropping, were then used to diversify and enrich the dataset, improving the model's training process to get better performance.
iv. The data was carefully annotated based on the insights and expertise of domain experts, adding essential information to facilitate the subsequent training of DL models.
    - Inter-annotator consistency was ensured using mean intersection over union (IoU) between annotators' masks.
    - Initial agreement of 0.97 between annotators is substantial agreement for models' training. For the rest of the cases with disagreements, structured adjudication process was followed.
    - Structured adjudication involved independent review by the third expert, without the knowledge of initial annotators.
v. The annotated data was then used to train the DL models. The models were created primarily to process, and classify potato crop disease and detect pests such as blight disease, leaf spot disease, leafroll virus (a visual symptom complex characterized by leaf rolling, independent of laboratory pathogen confirmation) [30, 31], and Colorado beetle in real time.
vi. The pre-trained YOLOv8 model was selected and fine-tuned on the custom collected dataset of diseased potato leaves and pests. Fine-tuning ensures appropriate fit of the model to specific object classes. DL models were able to provide accurate classification on potato crop disease and pests.

3.1 Image Acquisition

The first step of data preparation is image acquisition. Images of potato plants were captured directly in the field using a Logitech HD Pro C920 webcam (∼1 MP) at a native resolution of 1280× 720 pixels. The camera was mounted at a distance of approximately 33 cm above the leaf surface, providing close-up views that still contained contextual background (soil, neighboring plants). All images are in RGB color space and were stored in lossless/high-quality compressed format.

Data collection took place in Pakistan during both summer and winter cropping seasons. Images were collected at two sites:

i. The University Research Farm Koont, PMAS-Arid Agriculture University Rawalpindi (33.1166°N, 73.0111°E) and
ii. The University of Faisalabad, Okara campus (30.85°N, 73.53°E).

Across these sites, the dataset covers multiple growth stages (early vegetative, mid-vegetative, and pre-senescence) and seasonal windows (January-February and August-May). Importantly, no pesticides or fungicides were applied on the experimental fields to ensure that naturally occurring infections and infestations could be observed.

A total of 2,688 images were collected and excluding symptoms-free 285 images, led to a total of 2,403 images. Each image may contain one or more symptomatic regions or pests. The dataset includes four biotic stress categories: blight disease (late/early blight symptoms), leaf spot disease, leafroll virus, and Colorado beetle (larvae and adults). The "real-world field conditions" include substantial variability in illumination (direct sunlight, overcast, partial shadow), background clutter (soil, straw, weeds, irrigation pipes), leaf overlap, occlusion by neighboring plants, and motion blur due to wind.

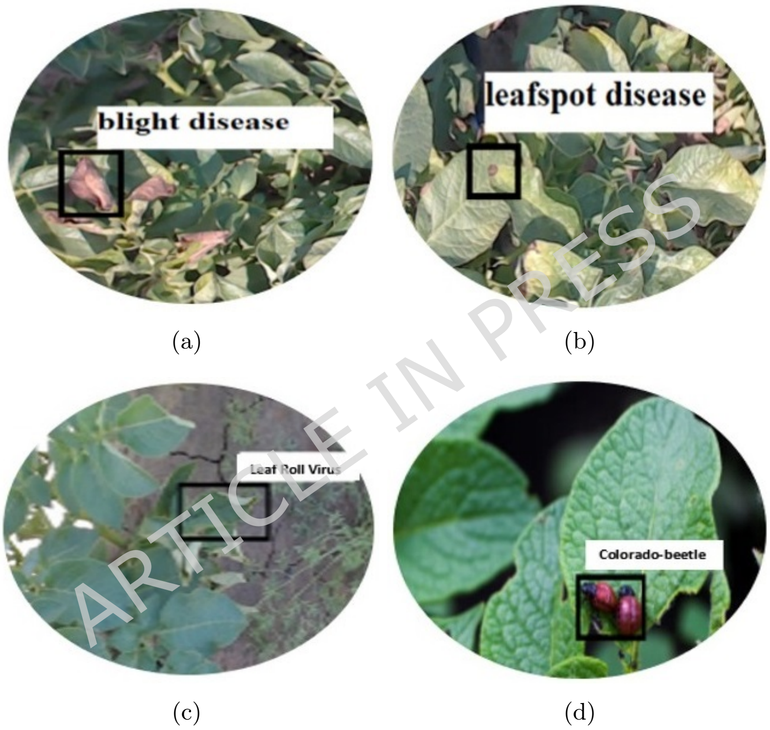


Fig. 2: Potato plant disease and pests classes.

Figure 2 illustrates representative samples for each class, while Figure 3 summarizes the temporal distribution of image collection. Note that in the earliest months after sowing (January in winter and August in summer), no visual disease or pest symptoms were present. Images from these symptom-free months were either (i) treated as healthy background samples or (ii) excluded from the present analysis, depending on the specific experiment; this point is clarified when discussing dataset composition and future extensions.

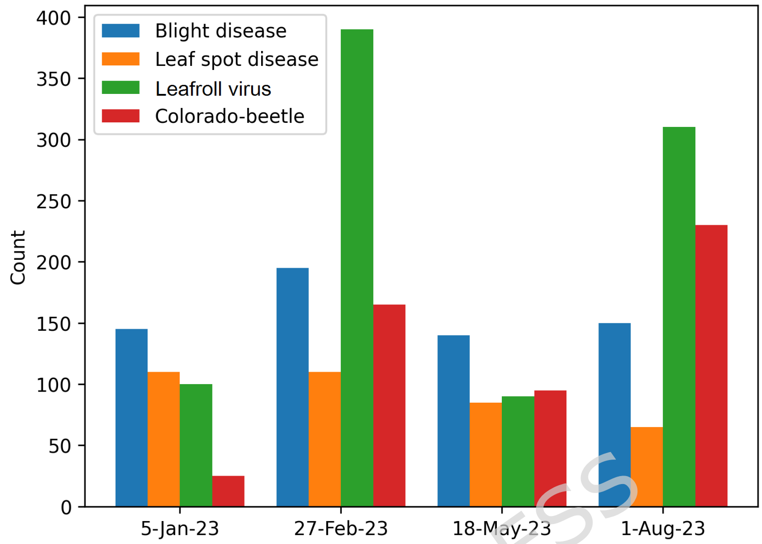


Fig. 3: Dataset description.

3.2 Dataset Preprocessing

Preprocessing is applied to enhance image quality and to ensure consistency across the dataset. These steps are crucial to eliminate noise and variations in the images [32]. Since the original images vary slightly in resolution and aspect ratio, images are resized to 640×640 pixels using bilinear interpolation. This size provides a good trade-off between detail preservation and computational efficiency for YOLO variants, and is consistent with recommended settings in prior studies on agricultural object detection [33]. Maintaining the same aspect ratio between the training and recognition images is critical for the YOLO model's best performance. A discrepancy in aspect ratio can lead to poor performance and shape distortion of detected objects [34].

The following preprocessing and cleaning steps were applied:

- **Data cleaning and normalization:** Images that were severely blurred, highly under- or over-exposed, or affected by strong camera artifacts were discarded. For the remaining samples, pixel intensities were scaled to [0, 1] and normalized using the mean and standard deviation of the training set.

- **Data augmentation:** To reduce overfitting and improve robustness to viewpoint and illumination changes, a set of geometric and photometric augmentations were implemented in Python (OpenCV + Albumentations). For each original training image, 2 to 4 augmented variants were sampled on-the-fly during training (so the effective number of augmented samples varies by epoch). The main transformations were:
  - Random horizontal and vertical flips with probability p = 0.5
  - Random rotations in the range [-70°, 70°] and 90° rotations
  - Random scaling in the range [0.8, 1.2]
  - Random cropping of up to 10-20% of the image while ensuring that at least 70% of each object remains visible.
  - Color jitter (brightness/contrast/saturation) and small Gaussian noise

  The probability of applying horizontal and vertical flips was set to p = 0.5. Rotation and scaling were applied with probability p = 0.5, while color jitter and Gaussian noise were applied with probability p = 0.3. All augmentation parameters were selected empirically based on validation performance to balance robustness and overfitting prevention.
  Augmentations were applied on-the-fly during training so that each epoch sees slightly different variants of the same base image, which is known to improve generalization.
- **Data annotation and quality control:** All images were annotated using the LabelImg tool in YOLO format (normalized bounding boxes and class IDs). Two annotators with expertise in plant pathology independently labeled the disease and pest regions. Ambiguous cases were resolved through discussion, and a third senior expert adjudicated disagreements. Only biotic stress symptoms (diseases and Colorado beetle) were annotated; images with predominantly abiotic stress (nutrient deficiency, water stress, mechanical damage) were excluded from the current dataset.

To ensure label quality, the following protocol was adopted:

- Each annotated image was cross-checked by at least one additional annotator.
- Samples with more than 50% overlap between different symptom types (e.g., overlapping lesions of distinct diseases) were carefully inspected. If experts could not reliably assign unique labels, those samples were discarded. Operationally, "overlap" was quantified as IoU > 0.5 between bounding boxes belonging to different classes.
- A small random subset (200 images) was re-annotated from scratch to estimate inter-annotator agreement.

This procedure resulted in a curated dataset with reliable bounding boxes and class labels suitable for training supervised detection models. Figure 4 shows resized images during data preprocessing.

Fig. 4: Image resizing.

The existing works made use of annotation tools like LabelImg, yolo_mark, roboflow, and others. The LabelImg is a widely used tool, and is publicly available on the Pipy website. To annotate the data for training YOLOv8-n, YOLOv8-m, YOLOv8-s, YOLOv7, Faster-RCNN, and YOLOv5 models, this study leverage LabelImg. Data was annotated using LabelImg with expert agricultural validation to ensure correctness. Only biotic stress samples were included, and co-occurring symptoms were carefully validated by experts before labeling. This ensured consistency and reduced misclassification. In addition, sample with more than 50% overlapping symptoms were discarded. Figure 5 shows an annotated rectangular image.

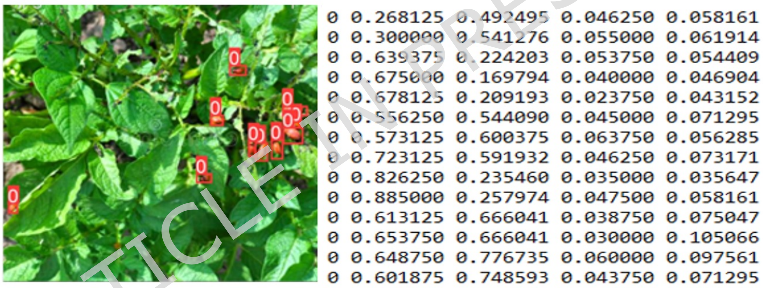


Fig. 5: Image annotation for YOLO model training using LabelImg.

3.3 Study Area

The data collection was carried out at two different locations: the university research farm Koont PMAS-AAUR. (coordinates 33.1166° N, 73.0111° E), and the University of Faisalabad Okara campus (coordinates 30.85° N, 73.53° E), both situated in the Punjab province of Pakistan, as depicted in Figure 6. The planting of the potato crop took place during two specific periods, January-February and August-May. The cultivation area covered approximately one acre, where no pesticides or fungicides were applied. The potatoes were sown with a line-to-line distance of 20cm and at a depth of 3-4 inches. Moreover, a substantial quantity of organic compost was spread on the ground before placing the potatoes into the soil.

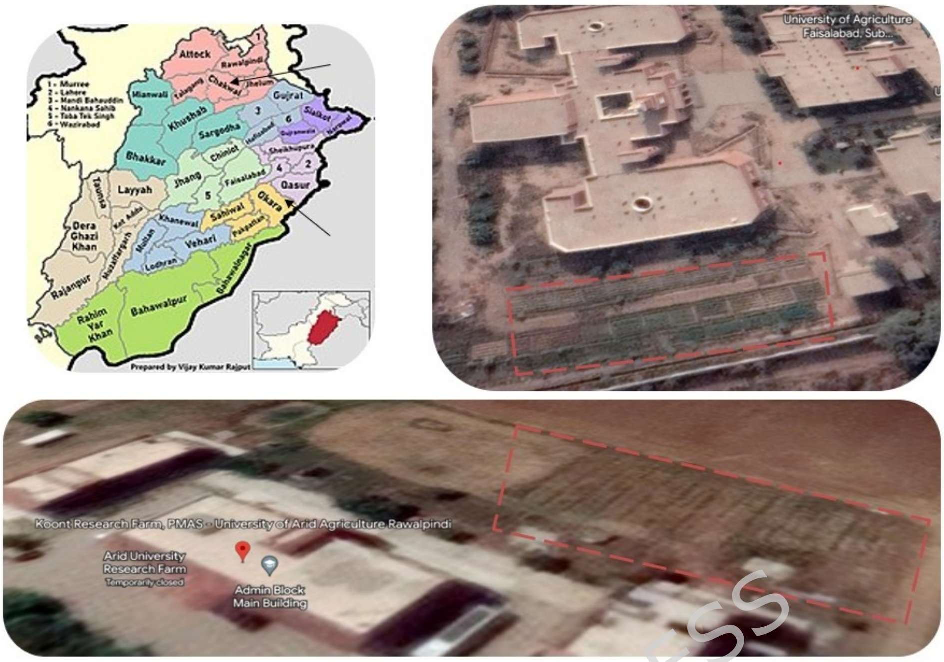


Fig. 6: Study area of potato crop. District-wise map of the Punjab, Pakistan (Top Left) [35], location for the 1st data collection site (Top Right by Google Earth) [36], and location for the 2nd data collection site (Bottom by Google Earth) [37].

3.4 Choice of YOLO Variants

This study evaluated different YOLO variants for potato plant disease and pest detection. The choice of the YOLO model is based on the following factors.

- YOLO variants are suitable for real-time object detection which makes them our first choice for disease detection. In addition, the inference time is better which makes them suitable for agricultural tasks involving the use of drones or other on-device applications.
- Existing studies reported the better performance of YOLO variants for plant disease detection [38]. YOLO-based models provide high accuracy and better mAP than other models and comparatively generalize well.
- YOLO models' performance can be enhanced with appropriate feature extraction and data fusion. These models adapt well for complex outdoor agricultural environments and provide robust results [39].
- Another characteristic feature of YOLO models is their capability to locate and identify multiple instances of disease in a single image. This feature is particularly suitable for scenarios, like ours, where multiple diseases are found on a single plant leaf [40].
- YOLO models balances speed and accuracy. Contrary to Faster-RCNN and similar other models with two-stage detectors, YOLO with its single-phase process, provides better inference time and accuracy.

3.5 Model Performance Evaluation

To enable reproducible comparisons across models, all detectors were trained and evaluated under a common protocol.

- **Train/validation/test splits:** The 2,403 images were partitioned at the image level into training (70%), validation (20%), and test (10%) sets using a stratified strategy so that the class distribution in each split approximated the global distribution. To reduce information leakage from near-duplicate field images (same plant/row/day), a manual and metadata-based check is performed to avoid placing visually near-identical frames across different splits; any suspected duplicates were kept within the same split. No images from the test set were used for model selection or hyperparameter tuning.
- **Initialization and fine-tuning:** All YOLO models (YOLOv8n/s/m, YOLOv7s, YOLOv5s/m) and Faster R-CNN were initialized from publicly available weights pre-trained on the MS-COCO dataset. The entire network is then fine-tuned on the potato dataset. This approach leverages generic features learned from large-scale object detection to compensate for the moderate size of the field dataset.
- **Hyperparameters:** Unless otherwise stated, this study used the default training hyperparameters provided in the official Ultralytics YOLOv5/YOLOv8 and YOLOv7 implementations. For clarity, the main settings include:
  - **Optimizer:** SGD with momentum 0.937.
  - **Initial learning rate:** LR_INIT (1e-3).
  - **Batch size:** 16 images per GPU.

- Number of epochs: 200 for YOLO-based models, 500 for YOLOv5s and 1,300 for Faster R-CNN.
  - **Early stopping:** training was stopped early if validation mAP did not improve for PATIENCE epochs.

  Hyperparameters were tuned based on the validation set using a small grid search around the defaults (e.g., varying the learning rate and batch size). The final values (reported above) were chosen to maximize validation mAP while keeping training stable.
- **Implementation and hardware:** All experiments were conducted on an Ubuntu 20.04 LTS system with an NVIDIA RTX A4000 GPU (32GB VRAM) and 16 GB system RAM. We used Python 3.8, CUDA 11.3, cuDNN 8.2, and OpenCV 4.2. YOLO implementations followed the official Ultralytics repositories.
- **Evaluation metrics:** This study reports mAP at IoU threshold 0.5 (mAP@0.5) as the primary metric, along with per-class AP, precision, recall, and training/validation loss curves. For completeness, we also report mAP@0.5:0.95 where appropriate. Inference time per image was measured as the average time to process a 640×640 image on the GPU (and optionally on CPU in an additional experiment). Besides a single train/validation/test split in this study, more robust statistical estimates including k-fold cross-validation, repeated runs with different random seeds, and ablation study is also carried out.
  In addition to k-fold validation, we repeated the YOLOv8-m training three times with different random seeds controlling weight initialization and data shuffling. The variation in final test mAP@0.5 was below ±0.4%, further confirming the stability of the reported performance.

4 Results and Discussions

The experiment in this study is designed to examine the efficacy of YOLO-based object detection in detecting disease and pests in potato plants, as well as the comparative findings under different models and the data augmentation impacts.

4.1 Experimental Setup

The entire experimental process, including training and validation, was carried out using the Ubuntu 20.04 LTS operating system, a Graphics Processing Unit, Nvidia RTX A4000. The NVIDIA RTX A4000 with 32 GB RAM is used on a Linux 64-bit OS. CUDA 11.3, cuDNN 8.2, Python 3.8, and OpenCV 4.2 are used in the experiment.

4.2 Model Training

The training, validation, and testing sets were produced by using 70%, 20%, and 10% of the images from the original data set, respectively. Both training and validation images used 640×640 dimensions. In this investigation, the original YOLOv-8 trained weight file was employed as the initialization parameter in the YOLOv-8 model. To avoid overfitting, several options were employed. By varying the weights and biases, the training process was visualized. It was used to dynamically monitor the model's training state and performance across iterations. Hyperparameter fine-tuning was carried out as well. In addition, early stopping criterion was used and the model's training was stopped when performance stopped improving. Dynamic monitoring was carried out during training and network's weight file was stored at each epoch to get the greatest detection performance. Specifically, overfitting prevention was assessed through the use of early stopping, k-fold cross-validation, and using random seeds, with performance monitored using validation loss, accuracy, and AUC metrics.

The result of the YOLOv8-m model is shown in Figure 1 in Supplementary Material. The model's parameters progressively increase during model iterations 0-70. As the number of iterations increased, the model's performance increased accordingly. The model index became stable after the 100th iteration, and the Precision value reached 0.89 and progressively stabilized. As a result, the model becomes stable and achieves 98.0% mAP between 90 and 200 iterations.

The loss function was used to calculate the YOLOv8-m predictor's performance in classifying, identifying, segmenting, and tracking the blight disease, leaf spot disease, leafroll virus, and Colorado beetle in a potato plant dataset (Figure 1 in Supplementary Material). It was used to evaluate the link between the data provided and the projected results. The model performs better when the loss is low, and vice versa. Each item contains two forms of loss: one for training loss and one for validation loss. The training loss for each epoch was calculated, and the testing loss for each epoch was determined.

4.3 YOLOv8-m Implementation

The essence of the YOLO target detection algorithm revolves around its compact size and efficient computational speed. YOLO is a cutting-edge approach for object detection, categorization, tracking, and segmentation. The YOLOv8-m model belongs to the YOLOv8 family and is recognized as the medium version due to its compact size and remarkable speed, while still delivering superior detection results. This model excels in object identification and classification, and it is complemented by instance segmentation and object tracking, making it a cutting-edge solution. Developed by Ultralytics, the same team behind the potent YOLOv5 model, YOLOv8-m incorporates various architectural updates and enhancements to further improve its performance. An essential characteristic of YOLOv8-m is its anchor-free design, where it directly predicts an object's center point within an image instead of relying on offsets from predefined anchor boxes.

Figure 1 in Supplementary Material, depicts the results of the selected model, which were used to view the identified potato diseases and pests, which included blight disease, leaf spot disease, leafroll virus, and Colorado beetle. In this study, unseen images were used to assess the model's viability, and the YOLOv8-m model effectively detected diseases and pests and gave a higher confidence score than other models. The threshold value was set to 30%, which implies that if the confidence score is more than or equal to it, the model will classify it. For identifying leaf spot disease, the model has a confidence score of 1.0, 0.9 for detecting leafroll virus disease, 0.9 for detecting blight disease, and 1.0 for detecting Colorado beetle pests. Figure 7 shows examples of disease detection.

Fig. 7: Example of effective disease detection in a potato plant using the YOLOv8-m model.

The precision and recall curves for each class are depicted in Figure 8. Among the three groups of potato plant disease and pests (blight disease, leaf spot disease, and leafroll virus), blight reaches the highest 0.98 mAP, leafroll virus earns the lowest 0.90 mAP, and leaf spot achieves a 0.98 mAP. The Colorado-beetle class received 0.97 mAP. Furthermore, the accuracy values of all classes are not significantly different. The model becomes stable after the 100th iteration and achieves above 90% outcomes.

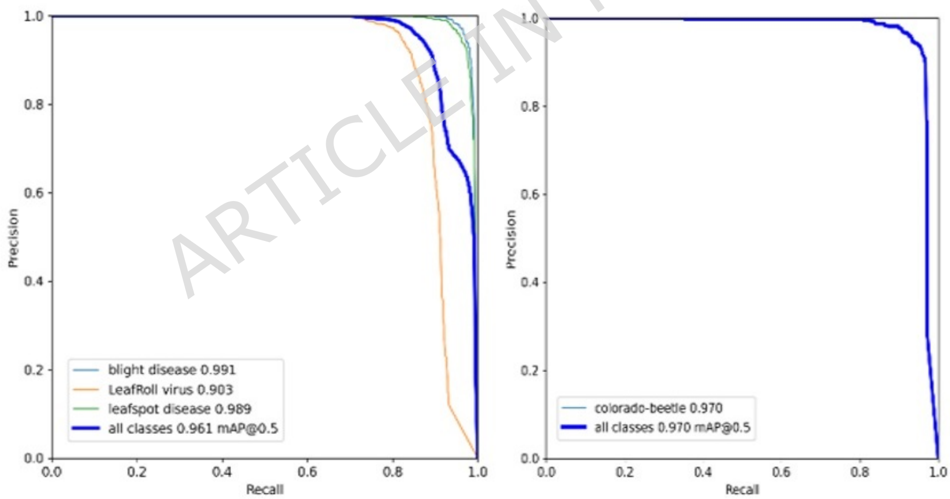


Fig. 8: YOLOv8-m model precision-recall curve of disease and pests.

4.4 YOLOv8-n Implementation

The YOLOv8-n is the smallest and fastest variant in YOLO series, while the YOLOv8 extra large (YOLOv8-x) is the largest but most accurate. After training the YOLOv8-n model, the precision and recall curves are depicted in Figure 2 in Supplementary Material. The model recall is 85% and the precision is 98%. The training loss is 2.7% and the validation loss is just 2.6%, which decreases when the training cycle increases. For identifying leaf spot disease, the model shows a confidence score of 1.0%, 0.9% for detecting leafroll virus disease, 0.9% for detecting blight disease, and 1.0% for detecting Colorado-beetle pests.

The precision and recall curves for each class are depicted in Figure 9. Among the three groups of potato plant disease and pests (blight disease, leaf spot disease, and leafroll virus), blight reaches the highest 0.98 mAP, leafroll virus earns the lowest 0.90 mAP, and leaf spot achieves 0.98 mAP. The Colorado-beetle class received 0.97 mAP. Furthermore, the accuracy values of all classes are not significantly different. The model becomes stable after the 100th iteration and achieves above 90% outcomes.

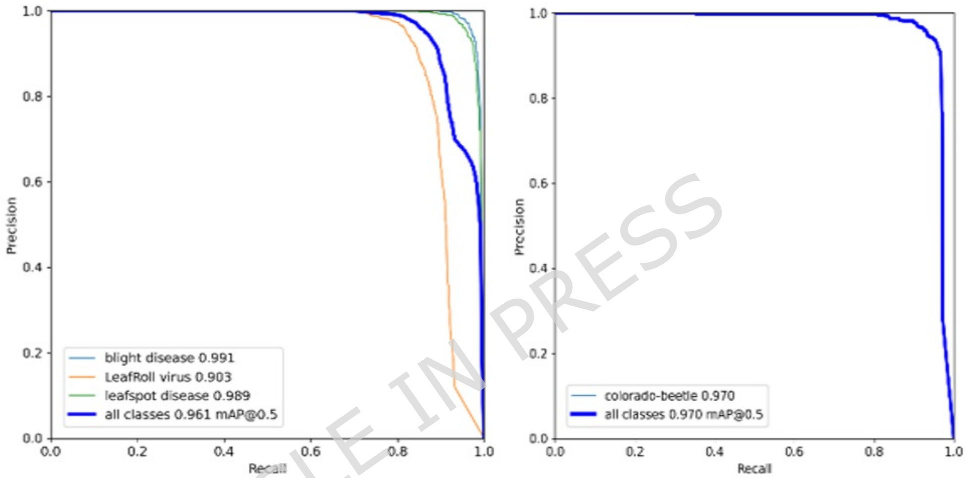


Fig. 9: YOLOv8-n model precision-recall curve of disease and pests.

4.5 YOLOv8-s Implementation

The YOLOv8 is the most recent version of the YOLO family, with improved features such as instance segmentation, pose/key points estimation, and classification, and the YOLOv8-s is also a version of it. The model recall is 82% and precision is 97%. The training loss is 2.7% and the validation loss is just 2.6%. The training loss is 3%, which is higher than nano but smaller than medium. It decreases gradually as the training cycle goes to its final output. Figure 3 in Supplementary Material, visualizes the YOLOv8-s performance.

4.6 YOLOv7-s Implementation

YOLOv7 is another YOLO family object detection model. YOLOv7 outperforms the best Cascade-Mask R-CNN models by 2% while inferring at a far faster rate. It is significant since prior R-CNN versions achieved substantially higher detection accuracy than single-stage detector designs. The model recall is 74% and the precision is 80%. The training loss is 2.7% and the validation loss is just 1.67%, and it decreases gradually when the training cycle increases.

4.7 Faster-RCNN Implementation

Faster R-CNN is an object identification model that improves on Fast R-CNN by combining the CNN model with a region proposal network. The RPN and the detection network exchange full-image convolutional features, allowing for almost cost-free region suggestions. It is a fully convolutional network that predicts object limits and objectness scores at each place at the same time. The visualization results of the model are shown in Figure 4 in Supplementary Material. After training the Faster-RCNN model, the accuracy curves are depicted in Figure 4 in Supplementary Material, which is 84%. Moreover, the average precision is 75%.

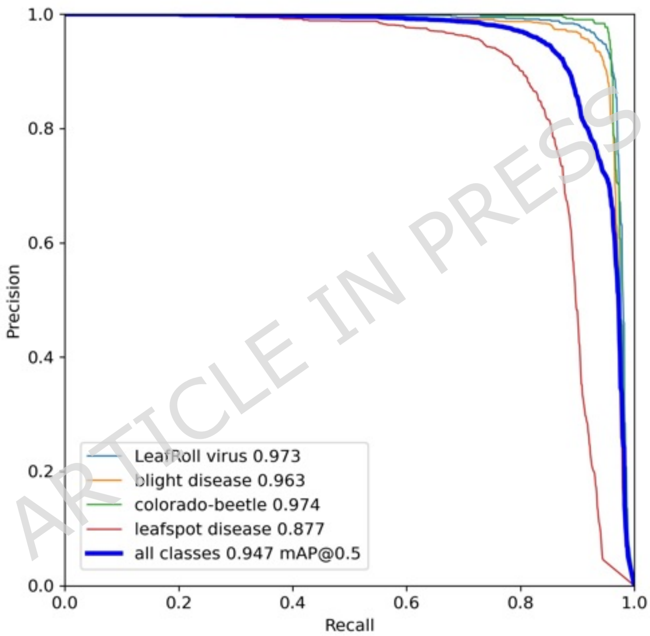


Fig. 10: YOLOv5-s model precision-recall curve of disease and pests.

4.8 YOLOv5 Implementation

To create image features, YOLOv5 employs a CNN backbone. These characteristics are merged in the model's neck and sent to the head. The collected characteristics are then interpreted by the model head to forecast the class of an image. YOLOv5 employs a novel approach for producing anchor boxes known as "dynamic anchor boxes." The ground truth bounding boxes are clustered using a clustering method, and the centroids of the clusters are used as anchor boxes. After training the YOLOv5-s model, the recall and confidence curves are depicted in Figure 10..

The model's precision-recall is 94% and recall is 97%. The model performance evaluations show that the training loss is almost 0.04% and the validation loss is also the same. The model shows stable output after 200 epochs, as shown in Figure 5 in Supplementary Material.

After training the YOLOv5-m model, the model's precision-recall is 100% and recall is 99%. The model performance evaluations show that the training loss is almost 0.04% and validation loss is 0.10%, which is shown in Figure 11. The model's training and validation performance is visualized in Figure 6 in Supplementary Material.

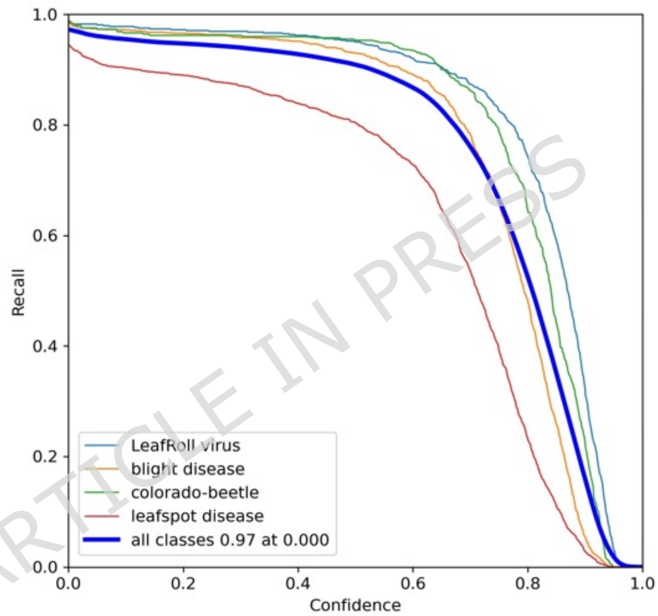


Fig. 11: YOLOv5m model precision-recall curve of disease and pests.

4.9 Comparison of the Applied Models

Table 1 shows the experimental results of the diseases: blight disease, leaf spot disease, and leafroll virus using different models such as YOLOv8-m, YOLOv8-n, YOLOv8-s, and YOLOv7s. As per the obtained results, the YOLOv8-m achieved the best blight (class A) performance with 99% average precision (AP), while the YOLOv8-s, YOLOv8-n, and YOLOv8-m performed best on blight disease (class A) and leafroll virus (class C) with 99%, 98%, and 98% AP, respectively. In terms of accuracy, the YOLOv8-m exceeded prior relevant versions of YOLO. Furthermore, this study looks at various performance assessors such as accuracy, recall, and inference time. According to the results, concerning the accuracy, the YOLOv8-m model yields a better value than YOLO variations, as well as a higher recall score.

The experimental findings of the pest Colorado beetle utilizing several models such as YOLOv8-m, YOLOv8-n, YOLOv8-s, and YOLOv7-s are shown in Table 2. Experimental results indicate that the YOLOv8-m achieves the best results, with a 97% AP. In terms of accuracy, the YOLOv8-m outperformed previous comparable YOLO versions as well. The YOLOv8-m model's accuracy is better than other YOLO variants, as well as a higher recall score. This study used YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv7-s, Faster-RCNN, and Yolov5-s to identify potato plant disease and pests in real-time and achieved 85%, 76%, 98%, 74%, 75%, and 87% accuracy, respectively.

Table 1: A comparison of the YOLO variants implemented in this research for disease detection.

Model Name	Average Precision			Img Size	Epochs	mAP@0.5	Precision%	Recall%	Inference Time
	A	**B**	**C**						
YOLOv8-n	98%	85%	97%	640×640	200	86	98%	65	7.5 ms
YOLOv8-s	98%	89%	98%	640×640	200	88	88%	87	7.7 ms
YOLOv8-m	99%	90%	98%	640×640	200	98	99%	88	7.2 ms
YOLOv7	82%	87%	80%	640×640	200	76	80%	74	10.1 ms

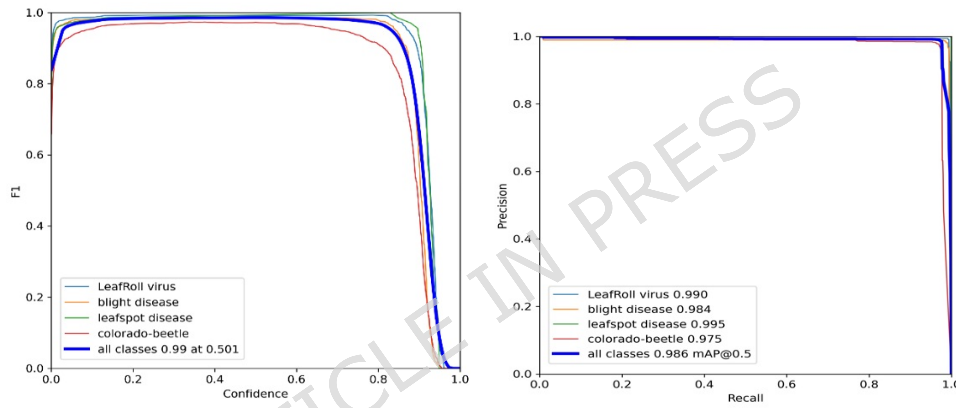
Table 2: Comparison of YOLO variants for Colorado Beetle Detection.

Model Name	Detection	AP (Colorado-beetle)	Img Size	Epochs	mAP@0.5	Precision %	Recall %	Inference Time
Yolov8-n	Pests	95%	640×640	200	85	86	87	12.1 ms
Yolov8-s	Pests	94%	640×640	200	76	80	76	7.6 ms
Yolov8-m	Pests	97%	640×640	200	97	98	88	7.4 ms
Yolov7	Pests	74%	640×640	200	75	80	74	10.1 ms

Table 3: A comparison of the models implemented in this research for disease and pests detection using 640×640 image size.

Model Name	Epochs	mAP@0.5	Inference Time
Yolov8-n	200	85	12.1 ms
Yolov8-s	200	76	7.6 ms
Yolov8-m	200	98	7.4 ms
Yolov7-s	200	75	12.1 ms
Yolov5-s	500	88	18.1 ms
Faster-RCNN	1300	75	6.43 ms

Tables 3 presents the results of various models concerning mAP and inference time. It shows that the best performance is obtained by YOLOv8-m using 200 epochs and the model achieves a 98% mAP. However, its inference time of 7.4 ms is little higher, compared to Faster-RCNN which has 6.43 ms inference time. The YOLOv5-s achieved an 88 mAP using 500 epochs and has the highest inference time of 18.1 ms. Results of YOLOv8-m are better than YOLOv8-n and YOLOv8-s concerning both the mAP and inference time. Figure 12 shows the F1 confidence curve for the Yolov8m, which shows that the F1 score is 99% and moreover the precision-recall curve is 98%.



(a) (b)

Fig. 12: Results for YOLOv8-m, **(a)** F1 score for detection, and **(b)** Precision-recall curve for detection.

4.10 Model Validation Using Additional Dataset

Table 4 shows the detection results of the models that gave the best training results. The used dataset is different than what is used in the current study for model training. The objective is to evaluate the generalizability of the YOLOv8-m for disease detection on the unseen data. The YOLOv8-m gave the best results for disease detection. It achieved 99.99% mAP using 1000 epochs, which is better than using 200 epochs. The Precision-Recall curve achieved by YOLOv8-m is 98%, which is higher than other models. Using an equal number of epochs for training, which is 200, YOLOv8-m yields better results including 97% precision and 85% recall. In addition, the number of false positives and false negatives are fewer for YOLOv8-m compared to YOLOv5-s.

Table 4: Cumulative accuracy metrics of YOLO at different training epochs for the automatic classification using 640×640 image size.

Model	Epochs	mAP@0.5	Batch	Precision %	Recall %	FP	FN
YOLOv5-s	200	60	16	65	60	108	87
	500	88	16	85	82	48	39
	1000	99.99	16	100	99	3	0
YOLOv8-m	200	98	8	97	85	40	7
	1000	99.99	16	99.99	99.99	3	3

4.11 Qualitative and Quantitative Error Analysis

To better understand the limitations of the evaluated models, qualitative and quantitative error analysis is also conducted on the YOLOv8-m predictions on the test set. Confusion matrices and per-class PR curves indicate that most errors arise from two scenarios:

i. **Confusion between visually similar diseases:** The most frequent misclassifications involve blight and leaf spot lesions when symptoms are at an early or mild stage, or when lesions co-occur on the same leaf. In such cases, lesion shape and color can partially overlap, leading to false positives of one disease class when another is present.
ii. **Small and partially occluded Colorado beetles:** Colorado beetle larvae and small adults under heavy foliage or in strong shadows are occasionally missed, especially when their projected size in the image is very small. This manifests as false negatives for the Colorado beetle class and slightly lower AP compared to leaf disease classes. Additional failure modes include:

- False positives on high-contrast but healthy leaf regions affected by specular reflections or mechanical damage.
- Reduced confidence under extreme illumination (very bright midday sun or deep shadow), where lesion boundaries appear washed out or noisy.
- Rare double detection (two overlapping boxes) for large lesions when the network interprets different parts of the same lesion as separate instances.

This analysis supports the need for more training data capturing small beetles, early-stage lesions, and extreme lighting, as well as complementary methods (e.g., multi-scale features, focal loss, or task-specific re-weighting) to further reduce false negatives in challenging scenarios.

The results of this study demonstrate that YOLOv8-m consistently outperforms the other evaluated models on the real-field dataset, achieving an mAP@0.5 of 98% on the test set. This performance gain can be attributed to several architectural improvements:

i. A deeper and wider backbone than YOLOv8-n and YOLOv8-s, enabling better representation of fine-grained texture differences between disease symptoms;
ii. An anchor-free head that directly regresses object centers and scales, which simplifies the optimization landscape compared to anchor-based methods; and
iii. A refined neck structure with improved feature aggregation across scales, which is particularly important for detecting Colorado beetles and small lesions at different resolutions.

Compared to older architectures such as YOLOv7, YOLOv5-s/m, and Faster R-CNN, YOLOv8-m benefits from more efficient feature reuse, improved training heuristics ("bag-of-freebies"), and stronger regularization, leading to higher accuracy at comparable, or lower inference time.

4.12 K-fold Validation Results

For additional verification, k-fold cross-validation is carried out using 5 folds for YOLOv8-m version, the best results are obtained using this model. Validation results are presented in Table 5. The results demonstrate stable and consistent performance across all folds, with minor variations attributable to data distribution differences. The high mAP@0.5 values confirm robust detection capability, while slightly lower recall indicates a conservative prediction strategy favoring precision.

To further quantify statistical robustness, the standard deviation and 95% confidence interval of the 5-fold cross-validation results was computed. The mean mAP@0.5 across folds was 97.8% with a standard deviation of 0.31%. The corresponding 95% confidence interval was 97.8% ± 0.27%, indicating low variance across folds and stable detection performance. These results confirm that the reported 98% mAP is not due to favorable data partitioning but reflects consistent model behavior.

Table 5: 5-fold cross-validation results using YOLOv8-m model.

Fold	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**
Fold 1	96.5	84.3	97.6
Fold 2	97.2	85.1	98.1
Fold 3	96.9	83.9	97.9
Fold 4	97.4	85.9	98.3
Fold 5	96.8	85.2	97.5
Average	**96.9**	**84.8**	**97.8**

4.13 Ablation Study

An ablation study is performed to further analyze the performance of Yolov8-m model for disease and pest detection. For this purpose, 50% , 30% and 20% data is used for training and model and results are presented in Table 6. The ablation study demonstrates that YOLOv8-m maintains strong detection performance even with reduced training samples. While recall decreases more rapidly than precision, the gradual decline in mAP confirms the robustness of the architecture under limited-data conditions.

Table 6: Results of ablation study for Yolov8-m model.

Training Data	mAP@0.5	Precision (%)	Recall (%)
50% data	93.2	92.3	77.5
30% data	89.7	88.9	72.1
20% data	86.1	85.6	68.4

4.14 Performance Comparison

Table 7 provides performance analysis with existing studies. The study [41] uses SSL for plant disease detection employing SSL with a vision Mamba encoder. Evaluation on PlantVillage, PlantDoc and Citrus datasets indicate 94.61%, 91.24%, and 89.83% accuracy, respectively. Similarly, [42] proposes PSMamba, a progressive SSL framework. The framework integrates teacher-student design with vision Mamba. Superior performance is reported on three different datasets, compared to existing approaches, with 98.97% on PlantVillage, 95.91% on PlantDoc and 92.22% on the Citrus dataset, respectively.

Direct numerical comparison should be interpreted cautiously, as prior studies primarily address classification on cropped or controlled images, whereas the method performs multi-object detection under real-world field conditions including background clutter, occlusion, and illumination variability.

Table 7: Comparison with recent state-of-the-art method.

Study	Task	Dataset Type	Metric	Performance
EfficientNet-ViT (2025) [11]	Classification	Cropped leaf images	Accuracy	85.06%
DenseNet201 TL (2024) [18]	Classification	Leaf images	Accuracy	92.5%
PotatoPestNet [12]	Pest classification	Controlled dataset	Accuracy	91%
StateSpace-SSL (2025) [41]	Disease detection	3 datasets	Accuracy	94.61%
PSMamba (2025) [42]	Disease detection	3 datasets	Accuracy	98.97%
YOLOv8 (2024)[43]	Detection	Lab/field mix	mAP	94-96%
Proposed YOLOv8m	Detection	Real-world field images	mAP@0.5	98%

4.15 Discussion

Experimental results show the superior performance of YOLOv8-m, compared to other DL models. The mAP of 98% highlights the effectiveness of the YOLOv8-m which offers a balance between computational efficiency and robustness. The depth of YOLOv8 offers subtle feature extraction for various disease patterns, leading to better output. Another important factor is the use of end-to-end object detection, contrary to existing works that focus on classification of pre-segmented leaf images. Background clutter and scale variation are additional challenges in the real-world environments. However, anchor-free prediction, and multi-scale feature aggregation makes YOLOv8 suitable to handle these challenges. The YOLOv8-m model shows the best results for blight disease (class A) with 99% AP. On the other hand, YOLOv8-s, YOLOv8-n, and YOLOv8-m also performed superb for blight (class A) and leafroll virus (class C), achieving 99%, 98%, and 98% AP, respectively. Overall, YOLOv8-m outperformed earlier YOLO versions in terms of accuracy, and other metrics.

YOLOv8-m performs better than YOLOv7, YOLOv5-s, and Faster R-CNN, which is primarily due to the architectural advancements incorporated in YOLOv8-m. Such enhancements enable the model to learn more efficient feature representations and achieve more accurate object localization. Moreover, YOLOv8m demonstrates better inference compared to the other models, highlighting its suitability for real-time deployment scenarios. When compared with recent frameworks, such as EfficientNet-ViT hybrid models, self-supervised feature extractors, and multimodal approaches like PotatoGPT, the proposed approach stands out with its focus on real-time and multi-object detection. While classification models often achieve high accuracy on cropped leaf images, they typically do not generate bounding boxes for lesion areas or pest instances. This limitation reduces their practicality for in-field crop monitoring and precision intervention tasks.

4.16 Limitations

Despite promising results, this study has several limitations that must be acknowledged:

i. The data used in this study was collected from two research farms in Punjab, Pakistan only. Consequently, the models learning might involve implicit region-specific patterns such as soil color, typical lighting, local characteristics, etc. Therefore, cross-regional validation and multi-site data collection must be used to confirm its generalizability.

ii. Although the dataset comprising 2,403 images and four biotic stress classes is real-filed data, it remains relatively small compared to large-scale object detection benchmarks. In addition, healthy plants are not included as an explicit class, and multiple pest species are not considered. Extending the dataset to additional classes, as well as increasing sample size is need for further investigation.

iii. For this study, only biotic stresses (diseases and Colorado beetle) were annotated. Abiotic stresses such as nutrient deficiencies, drought, salinity, and mechanical damage are not considered. Abiotic stresses can produce symptoms, similar to disease lesions, which makes detection and classification a challenging task. Future work should explicitly include abiotic symptoms and mixed-stress samples, and evaluate the model's ability to differentiate between them.

iv. The focus of this study YOLO-based detectors and Faster R-CNN, and several advanced approaches like Transformer-based detectors (DETR, ViTDet), and recent Mamba-style models, are not considered. Such models should be included in the investigation to analyze the performance difference of different design paradigms.

v. This study did not evaluate interpretability for the detection models. In practice, farmers and agronomists may benefit from visual explanations that highlight which regions of the leaf the network uses to decide on a particular disease or pest. Incorporating Grad-CAM, feature attribution, or attention visualization into the YOLOv8m pipeline is therefore an important extension.

vi. Inference times reported in this study are based on GPU measurement. Real-world deployments, however, may rely on CPU-only devices, smartphones, or low-power embedded systems. We did not conduct a systematic evaluation of latency and energy consumption on such hardware. Evaluating compressed or quantized versions of YOLOv8m and comparing them to lightweight models specifically designed for mobile deployment remains future work.

vii. This study follows image-level split for models which carries several limitations.

- Multiple images from the same plant with different angles, zoom, time points, etc. can lead to model memorizing plant-specific textures leading to model overfitting. It might lead to inflated accuracy, and precision.
- Similarly, images taken from the same row of plants might have similar lighting conditions, background, camera angles, etc., and model might not capture disease symptoms.
- Near-duplicate images involving burst captures or slightly shifted frames might produce very similar images. Following an image-level split might lead to an overstated performance due to similar images in training and testing sets.
- If early and late stages of infection from the same plant are separated across splits, the model may learn plant identity rather than pathology.

These limitations do not invalidate the main findings but highlight areas where additional experiments and data are needed to fully characterize and strengthen the proposed framework.

4.17 Future Work

In future work, we plan to:

- Extend the dataset to additional regions, cultivars, and management practices to better capture domain shift and improve generalizability.
- Incorporate abiotic stress classes and explicitly evaluate the model's ability to distinguish between biotic and abiotic symptoms.
- Explore self-supervised learning and domain adaptation techniques to leverage unlabeled field images and to adapt models trained in one region to new environments with minimal labeled data.

- Investigate Transformer-based and ensemble detectors (e.g., DETR, ViTDet, and CNN-ViT hybrids) as complementary baselines to YOLOv8, while maintaining real-time performance constraints.
- Integrate explainability tools (e.g., Grad-CAM) and interactive, multimodal interfaces similar to PotatoGPT to make model decisions more transparent and actionable for agronomists and farmers.
- Conduct deployment-oriented evaluations on mobile and embedded platforms, including model compression and quantization, to verify feasibility in resource-constrained settings.

By addressing these directions, we aim to move from a proof-of-concept detection framework toward a robust, scalable decision-support tool for sustainable potato production and, more broadly, for precision agriculture in smallholder and commercial systems.

5 Conclusion

A deep learning-based framework is evaluated in this study for potato diseases and pests detection using YOLOv8, YOLOv7, YOLOv5, and Faster-RCNN models. A custom real-field dataset from research farms in Pakistan was used for experiments. Results showed that YOLOv8-medium achieved the best performance, reaching 98% mean average precision. Its anchor-free design and deeper architecture enabled more accurate detection of fine-grained disease and pest features compared to smaller YOLO variants and older models. The proposed framework demonstrates that real-time, in-field monitoring of key potato biotic stresses (blight, leaf spot, leafroll virus, and Colorado beetle) is feasible with current hardware and algorithms. By providing accurate bounding boxes and class labels, the system could support early intervention, targeted pesticide application, and integration into precision agriculture workflows.

Despite the feasibility of real-time and infield monitoring of major potato biotic stresses, the results are limited to the specific region and growing conditions represented in the dataset. Cross-region and cross-season external validation was not performed, and therefore the model's generalizability beyond the studied environment cannot be assumed. The conclusions regarding deployment potential have been intentionally moderated to avoid overstatement.

For robustness and generalizability, future investigation should focus on using additional datasets from different agro-climatic regions across diverse conditions. Additionally, ensemble or stacking approaches may further enhance reliability by combining the strengths of multiple architectures. Field-level deployment studies, such as testing with drone imagery or mobile-based monitoring systems, are recommended to assess real-world performance. Expanding the system to include additional crop diseases and incorporating predictive analytics for disease progression may provide broader benefits for agricultural stakeholders.

Funding

This study was funded by the European University of Atlantic.

Conflict of interest

"The authors declare no conflict of interests."

Ethics approval and consent to participate

"Not applicable"

Consent for publication

"Not applicable"

Availability of data and materials

"The data can be requested from the corresponding authors."

Acknowledgments

"Not applicable"

Code availability

"Not applicable"

Clinical trial number

"Not applicable"

Authors' contributions

AA idea, formal analysis, writing original draft.
SUR data curation, idea, writing original draft.
KM methodology, data curation, formal analysis.
SGV funding acquisition, investigation, visualization.
LADL software, visualization, investigation.
AS project administration, software, methodology.
IA supervision, validation, writing - review and editing.
All authors read and approved the final manuscript.

References

[1] Afzaal, H., Farooque, A.A., Schumann, A.W., Hussain, N., McKenzie-Gopsill, A., Esau, T., Abbas, F., Acharya, B.: Detection of a potato disease (early blight) using artificial intelligence. Remote Sensing **13**(3), 1–17 (2021)

[2] Astani, M., Hasheminejad, M., Vaghefi, M.: A diverse ensemble classifier for tomato disease recognition. Computers and Electronics in Agriculture **198**(4), 107054 (2022)

[3] Upadhyay, N., Gupta, N.: Detecting fungi-affected multi-crop disease on heterogeneous region dataset using modified resnext approach. Environmental Monitoring and Assessment **196**(7), 610 (2024)

[4] Upadhyay, N., Sharma, D.K., Bhargava, A.: 3sw-net: A feature fusion network for semantic weed detection in precision agriculture. Food Analytical Methods, 1–17 (2025)

[5] Upadhyay, N., Bhargava, A.: Artificial intelligence in agriculture: applications, approaches, and adversities across pre-harvesting, harvesting, and post-harvesting phases. Iran Journal of Computer Science, 1–24 (2025)

[6] Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., Zou, X.: Plant disease recognition model based on improved yolov5. Agronomy **12**(2), 365 (2022)

[7] Upadhyay, N., Gupta, N.: Seglearner: A segmentation based approach for predicting disease severity in infected leaves. Multimedia Tools and Applications, 1–24 (2025)

[8] Coşkun, Ö.F., Aydın, A., Toprak, S., Başak, H.: Rootstock-mediated salinity resilience in cucumber (cucumis sativus l.): integrating physiological traits, genomic stability and machine learning. BMC Plant Biology (2025)

[9] Toprak, S., Coşkun, Ö.F.: Machine learning-based evaluation of nutrient distribution in grafted cucumber plants. PhytoTalks **2**(3), 457–466 (2025)

[10] Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B.: A review of yolo algorithm developments. Procedia Computer Science **199**, 1066–1073 (2021)

[11] Islam, M.N., Rahman, M.M., Islam, M.M., Ullah, M.A.: Performance evaluation of hybrid efficientnetv2b3 and vision transformer model for potato leaf disease detection. BMC Plant Biology **25**(1), 167 (2025)

[12] Kumar, A., Shukla, P.: Deep cnn feature extraction and wrapper-based svm classification for potato disease recognition. Revue d'Intelligence Artificielle **38**(2), 309–317 (2025)

[13] Ghosh, R., Sinha, S.: Transfer learning-based potato leaf disease detection using densenet201. arXiv preprint (2024)

[14] Ali, M., Khan, A., Shah, F., Zhang, Z.: A convolutional neural network-based framework for early and late blight detection in potato leaves. Artificial Intelligence Review **56**(4), 7823–7841 (2023)

[15] Kumar, S., Yadav, R., Sharma, V.: Potatopestnet: A novel cnn-based deep learning model for potato pest classification. arXiv preprint (2023)

[16] Iqbal, M., Khan, N.: Explainable ai with resnet for interpretable potato leaf disease detection. Frontiers in Artificial Intelligence **7**, 1449329 (2024)

[17] Abbas, A., Maqsood, U., Rehman, S., Mahmood, K., AlSaedi, T.: An artificial intelligence framework for disease detection in potato plants. Engineering, Technology & Applied Science Research **14**(1), 12628–12635 (2024)

[18] Karakan, A.: Potato beetle detection with real time and deep learning. Processes **12**(9), 2038 (2024)

[19] Zhu, H., Shi, W., Guo, X., Lyu, S., Yang, R., Han, Z.: Potato disease detection and prevention using multimodal ai and large language model. Computers and Electronics in Agriculture **229**, 109824 (2025)

[20] Poyato, D.H., Rull, J.D., Montes, R., Hernndez, I., Barrio, I., Echeverria, C.P., Tardaguila, J., Herrera, F., Herrera Poyatos: Small_data deep learning methodology for in field disease detection. arXiv preprint (2024)

[21] Sinamenye, J.H., Chatterjee, A., Shrestha, R.: Potato plant disease detection: leveraging hybrid deep learning models. BMC Plant Biology **25**(1), 647 (2025)

[22] Nishankar, S., Pavindran, V., Mithuran, T., Nimishan, S., Thuseethan, S., Sebastian, Y.: Vit-rot: Vision transformer-based robust framework for tomato leaf disease recognition. AgriEngineering **7**(6), 185 (2025)

[23] Astuti, T., Umar, A.N., Wahyudi, R., Rifai, Z.: Enhancing potato leaf disease detection: Implementation of convolutional vision transformers with synthetic datasets from stable diffusion. JOIV: International Journal on Informatics Visualization **8**(4), 2054–2065 (2024)

[24] Zhang, H., Li, S., Xie, J., Chen, Z., Chen, J., Guo, J.: Vmamba for plant leaf disease identification: design and experiment. Frontiers in Plant Science **16**, 1515021 (2025)

[25] Yuan, C., Li, S., Wang, K., Liu, Q., Li, W., Zhao, W., Guo, G., Wei, L.: Mamba-yolo-ml: A state-space model-based approach for mulberry leaf disease detection. Plants **14**(13), 2084 (2025)

[26] Ilsever, M., Baz, I.: Consistency regularization based semi-supervised plant disease recognition. Smart Agricultural Technology **9**, 100613 (2024)

[27] Mu, J., Feng, Q., Yang, J., Zhang, J., Yang, S.: Few-shot disease recognition algorithm based on supervised contrastive learning. Frontiers in plant science **15**, 1341831 (2024)

[28] Faria, F.T.J., Moin, M.B., Alam, M.S., Wase, A.A., Sani, M.R., Hasib, K.M.: Potatogans: Utilizing generative adversarial networks, instance segmentation, and explainable ai for enhanced potato disease identification and classification. arXiv preprint arXiv:2405.07332 (2024)

[29] Nagarathinam, T., Venkatesan, M.T., Arulmozhi, K.: A unified survey of supervised, unsupervised, and semi-supervised learning techniques for plant leaf disease detection

[30] Shattock, R.: Compendium of potato diseases, WR Stevenson. Wiley Online Library (2002)

[31] Taliansky, M., Mayo, M.A., Barker, H.: Potato leafroll virus: a classic pathogen shows some new tricks. Molecular plant pathology **4**(2), 81–89 (2003)

[32] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., Wang, K.: Image preprocessing in classification and identification of diabetic eye diseases. Data Science and Engineering **6**(4), 455–471 (2021)

[33] Sharma, P., Hans, P., Gupta, S.C.: Classification of plant leaf diseases using machine learning and image preprocessing techniques. In: Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering, vol. 20, pp. 480–484 (2020)

[34] Jeong, H.J., Park, K.S., Ha, Y.G.: Image preprocessing for efficient training of yolo deep learning networks. In: 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 635–637 (2018)

[35] Government of Pakistan: District-wise map of the Punjab, Pakistan. https://www.surveyofpakistan.gov.pk/SiteImage/Downloads/province_map_punjab.pdf

[36] Google Earth: University of Agriculture Faisalabad, Okara Campus, Pakistan. https://earth.google.com/web/search/university+of+agricultureokara+campus/@30.8510545,73.5330093,186.79655602a,713.32702128d,35y,0h,0t,0r/data=CpUBGmcSYQolMHgzOTE4MGJkMDBmMDNmZmZmOjB4MmRlNWEwYzJiZGZiYWIyNBkXnSy13tk-QCGw_QnTHGJSQComdW5pdmVyc2l0eSBvZiBhZ3JpY3VsdHVyZSBva2FyYSBjYW1wdXMYAiABIiYKJAkaeJEa7x82QBEXeJEa7x82wBn7cCN0T-NJQCE0DRqS3VRKwEICCAE6AwoBMEICCABKDQj__________8BEAA?authuser=0

[37] Google Earth: Koont Research Farm, PMAS-University of Arid Agriculture Rawalpindi, Pakistan. https://earth.google.com/web/search/Koont+Research+Farm,+PMAS+-+University+of+Arid+Agriculture+Rawalpindi,+Chakwal+Rd,+Mandra,+Pakistan/@33.1166308,73.0111446,527.00020069a,694.15106496d,35y,0h,0t,0r/data=CtQBGqUBEp4BCiUweDM4ZGZmN2Q4NzRjZjhkZDk6MHhlN2U5ZWJiNmFkZmQ2OTNhGSdHpujtjkBAIQgbQ6W2QFJAKmNLb29udCBSZXNlYXJjaCBGYXJtLCBQTUFTIC0gVW5pdmVyc2l0eSBvZiBBcml

kIEFncmljdWx0dXJlIFJhd2FscGluZGksIENoYWt3YWwgUmQsIE1hbmRyY
SwgUGFraXN0YW4YAiABIiYKJAk9sqT5hdo-QBENFWZsN9k-QBlBg-OFe2J
SQCEieDAgvmFSQEICCAE6AwoBMEICCABKDQj__________8BEAA?authuse
r=0

[38] Kaur, R., Mittal, U., Wadhawan, A., Almogren, A., Singla, J., Bharany, S., Hussen, S., Rehman, A.U., Al-Huqail, A.A.: Yolo-leafnet: a robust deep learning framework for multispecies plant disease detection with data augmentation. Scientific Reports **15**(1), 28513 (2025)

[39] Qin, R., Wang, Y., Liu, X., Yu, H.: Advancing precision agriculture with deep learning enhanced sis-yolov8 for solanaceae crop monitoring. Frontiers in Plant Science **15**, 1485903 (2025)

[40] Li, D., Ahmed, F., Wu, N., Sethi, A.I.: Yolo-jd: A deep learning network for jute diseases and pests detection from images. Plants **11**(7), 937 (2022)

[41] Mamun, A.A., Zhang, M., Ahmedt-Aristizabal, D., Hayder, Z., Awrangjeb, M.: Statespace-ssl: Linear-time self-supervised learning for plant disease detection. arXiv preprint arXiv:2512.09492 (2025)

[42] Mamun, A.A., Zhang, M., Ahmedt-Aristizabal, D., Hayder, Z., Awrangjeb, M.: Psmamba: Progressive self-supervised vision mamba for plant disease recognition. arXiv preprint arXiv:2512.14309 (2025)

[43] Abid, M.S.Z., Jahan, B., Al Mamun, A., Hossen, M.J., Mazumder, S.H.: Bangladeshi crops leaf disease detection using yolov8. Heliyon **10**(18) (2024)