



RESEARCH ARTICLE

Hybrid deep segmentation architecture using dual attention U-Net and Mask-RCNN for accurate detection of pests, diseases, and weeds in crops

Deepa Ramachandran VR^{1*}, Kamalraj N²

Abstract

Early and accurate identification of pests, diseases, and weeds in modern agriculture is crucial for sustainable crop management and yield optimization to increase productivity. This research proposes a hybrid deep segmentation framework that integrates dual attention UNet and Mask-RCNN methods to enhance the precision and reliability of plant disease detection under diverse environmental conditions. The core objective is to improve segmentation accuracy and object localization, particularly in complex field imagery with overlapping foliage, variable lighting, and background noise. The proposed architecture uses the plant-village dataset, which includes a diverse collection of annotated crop images representing multiple classes of pests, diseases, and weed species. The dual attention UNet emphasizes salient spatial and channel-wise features, enabling refined pixel-level segmentation of affected regions. This is followed by a Mask-RCNN module that performs instance-aware segmentation and bounding box localization, facilitating detailed identification of individual anomalies even in cluttered scenes. The framework is further enhanced through data augmentation and transfer learning strategies to support generalization across varying crop types. Experimental evaluation reveals that the proposed deep learning-based model achieves a detection accuracy (DA) of 96.5%, an F1-score of 95.2%, AUC-PR of 97.4%, sensitivity of 96.5%, scalability of 96.2% and a processing time (PT) of 12 seconds per batch, demonstrating both precision and efficiency. Moreover, the architecture shows a scalability of 96.8%, ensuring robustness in large-scale deployments. The comprehensive results are compared with baseline models such as CNN, faster R-CNN, and CBAM. The hybrid integration of instance-aware detection and attention-driven segmentation, explicitly designed for agricultural situations, shows the novelty, and the model improves detection quality by capturing fine-grained spatial characteristics and allowing for the thorough separation of overlapping anomalies compared to traditional CNN-YOLO pipelines. This model presents a reliable solution for real-time smart agriculture systems aimed at proactive crop health management.

Keywords: Attention mechanism, Deep learning, Mask-RCNN, Plant-village dataset, Smart agriculture, U-Net model.

¹Research Scholar, Department of Computer Science, Park's College (Autonomous), Chinnakkarai, Tirupur, Tamilnadu, India - 641605

²Vice Principal, Park's College (Autonomous), Chinnakkarai, Tirupur, Tamilnadu, India - 641605

***Corresponding Author:** Deepa Ramachandran VR, Research Scholar, Department of Computer Science, Park's College (Autonomous), Chinnakkarai, Tirupur, Tamilnadu, India - 641605, E-Mail: dpasudi@gmail.com

How to cite this article: Ramachandran, D.V.R., Kamalraj, N. (2025). Hybrid deep segmentation architecture using dual attention U-Net and Mask-RCNN for accurate detection of pests, diseases, and weeds in crops. *TheScientificTemper*, **16**(7):4521-4534. Doi: 10.58414/SCIENTIFICTEMPER.2025.16.7.04

Source of support: Nil

Conflict of interest: None.

Introduction

In recent years, the agricultural sector has grown dependent on intelligent technologies to manage crop health, productivity, and sustainability challenges. Among these challenges, the early and accurate detection of pests, diseases, and weeds remains a critical concern, especially in large-scale farming environments where manual inspection is labor-intensive and error-prone. As plant health anomalies significantly reduce yield and quality, precision-based detection mechanisms using deep learning have emerged as transformative tools in modern agricultural technology solutions. State-of-the-art models like deep learning-based dual-CNN and faster R-CNN have demonstrated significant promise in feature extraction and object detection tasks. Dual CNN architectures enhance representation learning by capturing complementary feature hierarchies through

parallel convolutional streams, while faster R-CNN offers robust region proposal and classification for real-time object localization. In parallel, the convolutional block attention module (DL-CBAM) has been integrated into deep architectures to refine spatial and channel-wise attention, which significantly improves the attention on critical regions in complex agricultural images. Despite these advancements, challenges include overlapping foliage, inconsistent lighting, visual similarity between classes, and dense background that continue to slow down the segmentation precision and overall detection reliability. This research gap underscores the need for a more integrated approach that identifies anomalies and segments them with high spatial resolution.

Recent research highlights significant strides in agricultural image analysis through deep learning. A novel U-Net hybrid proposed for pest segmentation demonstrated improved boundary clarity and class separation in noisy images (N. Biradar *et al.*, 2024). Image segmentation methods for crop health assessment continue to evolve, with attention mechanisms and multi-scale architectures improving pixel-level accuracy (L. Lei *et al.*, 2024). Deep learning-driven weed–crop differentiation is also gaining traction for real-time field automation (H.-R. Qu *et al.*, 2024). A comprehensive review emphasized the superiority of DL models over traditional techniques in identifying plant diseases and pest infestations across complex conditions (M. Shoaib *et al.*, 2025). This research proposes a novel hybrid segmentation framework that synergistically combines Dual Attention UNet with Mask-RCNN to address these limitations. The dual attention UNet component focuses on extracting spatially enhanced feature maps by integrating channel and position attention, while Mask-RCNN performs instance-aware segmentation and precise boundary localization. This dual-stage system is particularly suited for agricultural scenarios where multiple pests or diseases coexist within the same frame. The study targets the explicit detection under varying environmental conditions using the Plant-Village dataset, a widely used benchmark for plant anomaly detection. By addressing the limitations of existing pipelines, this work contributes a highly scalable, interpretable, and accurate solution to advance smart agriculture toward real-time, field-deployable systems.

Problem Statement

To develop a robust hybrid deep learning framework that integrates dual attention UNet and Mask-RCNN for accurately detecting and segmenting pests, diseases, and weeds in crops. The primary goal is to enhance pixel-level segmentation accuracy and instance-level detection precision under real-world agricultural conditions, where image complexity, overlapping anomalies, and environmental variability often degrade conventional model performance. The proposed hybrid model employs

the Dual Attention method to highlight spatial and relevant intrinsic features during segmentation, which improves the model's focus on core affected regions. Simultaneously, using instance-aware segmentation, Mask-RCNN is used to localize and differentiate multiple anomalies within a single frame. This dual-stage architecture ensures that even subtle or co-occurring threats are effectively segmented and detected. The significant contributions include the design of an integrated attention-guided pipeline, using the Plant-Village dataset for diverse crop anomaly training, and optimizing processing time and scalability for field deployment. By combining the strengths of attention-based feature enhancement with region-specific detection, the framework directly addresses existing challenges in misclassification, low boundary clarity, and poor generalization across crop varieties.

Objectives of the Proposed Work

The main objective of this research is to introduce a robust & intelligent deep learning model for accurate detection and segmentation of pests, diseases, and weeds in crops under complex agricultural environments. The key objectives are i) integration of dual attention UNet for precise pixel-wise segmentation; ii) Mask-RCNN to achieve instance-level localization of multiple coexisting anomalies; iii) enhancing spatial attention mechanisms to improve boundary clarity and reduce misclassification; iv) to ensure high adaptability and scalability across varying crop types and environmental conditions using the Plant-Village dataset; v) to reduce system processing time and maintains high detection accuracy, making it highly suitable for real-time usage in precision farming; etc. To attain the objectives, the major steps are followed:

- **Preprocessing:** Curate and preprocess agricultural image data from the Plant-Village dataset.
- **Segmentation:** Train Dual Attention UNet for detailed spatial and channel-aware segmentation.
- **Localization:** Integrate Mask-RCNN for instance-aware object detection and boundary refinement.
- **Data Augmentation:** Optimize the combined architecture using data augmentation and transfer learning.
- **Evaluation:** Evaluate model performance using accuracy, F1-score, sensitivity, AUC-PR, and scalability metrics.

Related Works

A recent advance in deep learning has significantly transformed automated agricultural diagnostics, especially in pest, disease, and weed detection systems. In a comprehensive benchmarking study, a multi-stage Faster R-CNN-based framework was developed that accurately classified and detected individual weed species at the object level. The model highlighted the strength of region-based

CNNs in delineating multiple overlapping plant types in field environments (Hasan *et al.*, 2023). Focusing on real-time edge computing, a lightweight YOLOv8 model is designed and enhanced with the multi-attention module (CBAM), enabling early cotton weed detection on edge devices. The integration of CBAM improved feature refinement and allowed the model to dynamically prioritize critical visual cues, enhancing detection under resource-limited conditions (Karim *et al.*, 2024). In another study, a dual CNN framework was proposed to detect pests, diseases, and weeds. The D-CNN multi-path convolutional design facilitated enriched feature learning, which proved vital for improving crop yield in dense vegetation imagery (Meena *et al.*, 2023). YOLOv5 was used to identify multiple weed growth stages in wheat fields. By leveraging stage-wise classification, the YOLOv5 system achieved high detection accuracy across phenological phases, reinforcing the applicability of deep detection models across dynamic agricultural timelines (Guzel *et al.*, 2024). Small pest detection using deep learning object detection algorithms was explored, and the importance of detecting minor infestations before escalation was highlighted, using fine-grained annotation strategies in real-time surveillance models (Khalid *et al.*, 2023). An enhanced deep learning feature extraction model specifically optimized for pest and disease detection was introduced, where the model uses integrated visual feature amplification strategies to improve detection precision and reduce false alarms across varied backgrounds (Yuan *et al.*, 2024). Table 1 shows the literature analysis of prevailing models.

Research Gap

Though deep learning applications have shown significant progress in precision agriculture, most existing models struggle with precise segmentation and detection when multiple anomalies, such as pests, diseases, and weeds, co-occur within a single image. Traditional CNN and YOLO-based approaches often lack the spatial awareness to isolate overlapping regions or small-scale infestations in complex field environments. Furthermore, current models rarely integrate attention mechanisms and instance-level detection in a unified framework. Limited research on attention-guided segmentation specifically tuned for agricultural datasets like plant-village. This creates a clear gap for a robust, dual-stage architecture that enhances feature focus through attention and also segments and localizes anomalies with high precision. Addressing this gap improves early crop health diagnostics and enables smarter, real-time agricultural decision-making.

Proposed Methodology

The proposed framework combines dual attention U-Net with Mask-RCNN (DA U-Net with Mask-RCNN) to deliver high-precision detection and segmentation of crop pests, diseases, and weeds. First, we preprocess the plant-village

dataset, which includes thousands of annotated images from multiple crop species under varied lighting and growth stages. Major features extracted include color texture, leaf shape patterns, disease lesion boundaries, and degree of infestation, providing rich representations for model learning. Dual attention U-Net processes these features through channel attention, highlighting critical spectral cues (e.g., discolorations), and spatial attention, sharpening focus on damaged areas. This attention-guided path produces refined pixel-level segmentation masks, delineating anomalies amidst overlapping foliage. These segmentation outputs are then fed into Mask-RCNN, which performs instance-level detection, generating bounding boxes and per-instance masks for individual pests, diseases, or weeds. This dual-stage approach ensures holistic segmentation and precise object localization within cluttered scenes. Applied data augmentation (rotations, color shifts) and transfer learning using ImageNet-pretrained backbones to boost accuracy further. This integrated deep learning model is fully optimized, which proves segmentation accuracy, box regression, mask quality, and attention consistency. This dual attention U-Net with Mask-RCNN model leads to robust, interpretable, and scalable detection performance across diverse crop conditions.

Materials and Methods for Implementation

To assess the effectiveness of the proposed U-Net with Mask-RCNN dual attention model, a structured implementation approach was adopted, combining multi-task deep learning models, standardized datasets, with robust training practices. The focus was on achieving high accuracy in detecting pests, diseases, and weeds under diverse agricultural conditions. This section outlines the detailed materials and methods, such as the dataset employed, the model architecture, implementation tools, training configurations, and optimization strategies used throughout the development process. Specific attention was given to data preparation, model scalability, and generalization to ensure the system performs reliably in real-world applications.

- **Dataset Used:** The Plant-Village dataset was used, containing annotated images of healthy and affected crop leaves with pests, diseases, and weeds.
- **Framework:** A hybrid architecture was developed by integrating Dual Attention UNet for segmentation and Mask-RCNN for instance-level detection.
- **Implementation Tool:** Python, with TensorFlow and MATLAB, is used for implementation, computation and comparative analysis.
- **Training and Testing Ratio:** The dataset is split in a ratio of 80% training and 20% testing to ensure proper evaluation of learning effectiveness.
- **Data Augmentation:** Adequate techniques such as rotation, flipping, color jittering, and random cropping were applied to improve model generalization.

Table 1: Literature analysis of specific state of art object detection methods

<i>Authors</i>	<i>Methods Adopted</i>	<i>Merits</i>	<i>Limitations</i>
Buckner <i>et al.</i> (2021)	High-throughput image segmentation with machine learning	Scalable across multi-resolution plant imaging; cross-scale adaptability	Less optimized for real-time performance; lacks fine-grain object separation
Demilie (2024)	Comparative study of classical ML vs. DL techniques	Provides statistical validation of model performances	Limited use of advanced segmentation; lacks real-time experimentation
Devi <i>et al.</i> (2024)	GAN-based secure data routing for IoT	Introduces bio-inspired robustness and data loss minimization	Focuses on data transmission; where image classification is not done
Dinesh & Lakshmanan (2025)	Multiclass semantic segmentation on citrus leaves	Identifies severity levels along with disease classification	Dataset focused on citrus plants only; model generalization not addressed
Dolatabadian <i>et al.</i> (2025)	Image-based disease detection using ML classifiers	Large-scale dataset and robust classification strategies	Less use of attention-based deep learning models
Eldho & Nithyanandh (2024)	3D CNN model on CT-DICOM images	Accurate volumetric analysis; robust segmentation for medical datasets	Deep object detection is don, but lacks highlighted masking for classifictaion
Faisal <i>et al.</i> (2025)	Customized deep learning for cotton disease detection	High precision detection under diverse environmental settings	Model performance in multiclass settings is not validated
Hasan <i>et al.</i> (2023)	Extended KDE clustering for coffee leaf diseases	Novel unsupervised learning for disease classification	Limited interpretability in overlapping disease classes
Kim & Park (2022)	MTS-CNN (multi-task semantic CNN)	Efficient in distinguishing weeds and crops simultaneously	Higher computational complexity in dense scenarios
Liu & Wang (2021)	Review of DL methods for plant disease detection	Broad coverage of DL architectures and datasets	Deep segmentation and masking of localized features
Nithyanandh (2025)	YOLOv8 with Deep CNN on soft computing-based object detection	High speed detection with efficient feature mapping	Focused more on system design than object detection and deployment results
Nithyanandh <i>et al.</i> (2023)	Firefly algorithm based EAP-IFBA with adaptive sleep scheduling	Energy-efficient and secure IoT-based communication	Focused more on IoT image-based agricultural detection
Prabhu <i>et al.</i> (2025)	Bio-inspired secure routing in 6G wireless agriculture systems	Enhanced authentication and energy optimization	Sensor optimization is focused, lacks deep segmentation and masking
Selvam & Joy (2024)	AEN + Mask R-CNN with multi-variable feature selection (custom PL dataset)	Improved lesion detection with precise segmentation	Lacks multiclass scalability for large datasets
Silva <i>et al.</i> (2024)	Weed segmentation using UAV images and deep learning	Real-time UAV-based monitoring; efficient for large field areas	Weather and drone limitations affect consistency
Singh <i>et al.</i> (2020)	Plant-Village dataset creation for plant disease detection benchmarking	Open-source benchmark with class variety and annotation	Imbalanced class distribution in dataset
Upadhyay <i>et al.</i> (2025)	Review on DL models, datasets and trends in precision agriculture	Summarizes models with detailed insights on future directions	Scalability issues on multiclass image dataset
Arularasan <i>et al.</i> (2024)	Deep learning-based sign language recognition with visual feature extraction	Effective gesture modeling with deep architecture integration	Object detection with visualized feature extraction is focused, lacks deep spot localization

- **Transfer Learning:** Pretrained weights (ImageNet) were used to initialize backbone layers for faster convergence and improved accuracy.
- **Training Configuration:** Model trained for 120 epochs using Adam optimizer, monitored with performance metrics like F1-Score, Detection Speed, and AUC-PR.

PLANT-VILLAGE Dataset Attainment and Preprocessing

The proposed U-Net with Mask-RCNN model uses the Plant-Village dataset, which was meticulously built to support realistic disease detection of pests and weeds in crops. It contains 54,303 images across 38 categories,

with healthy and diseased leaves from 14 crop species. To ensure robustness, we adopted the TensorFlow Datasets version in this research, which maintains consistent image quality and standardized labeling. Before the model training, images are resized to 224×224 pixels to align with standard CNN input dimensions. Pixel values were then normalized per channel using the formula. $x_{norm} = \frac{x - \mu}{\sigma}$, where μ and σ are the mean and standard deviation, which are calculated over the training set. This normalization ensures stable convergence across varying lighting conditions. Additionally, we compared RGB, HSV, and lab color spaces to emphasize contrasts between healthy tissue and lesions. Augmentation techniques such as random rotation ($\pm 20^\circ$), horizontal and vertical imaging flips, brightness & contrast adjustments, and Gaussian noise are used to enhance training variety. This process expanded the dataset by 5x and improved generalization to real-world scenarios. Self-contained class balancing ensured that all 38 categories were represented equitably during batch training. Bounding box annotations were derived from image-level labels and used only for diagnostic evaluation, reserving mask-based labels for segmentation tasks. Altogether, this preprocessing pipeline standardized inputs, mitigated illumination bias, and enhanced feature separation, setting the stage for effective attention-based segmentation and instance detection. Table 2 shows the Plant-Village Crop Species and Disease Classes to spot the classification.

To ensure balanced learning and reliable validation, the complete Plant-Village dataset is split into an 80:20 ratio, where 80% is used for training and 20% reserved for testing and validation purposes. In order to preserve class

distribution across both sets stratified sampling method was applied to ensure that the minority classes receive sufficient representation during learning, while the model’s generalization ability is assessed on hidden samples, which reflects the real-world variability and intra-class differences.

Key features include leaf texture patterns, which help distinguish between healthy and infected surfaces, and color distortions, such as yellowing, browning, or lesion-specific discolorations that signal early disease onset. Edge irregularities and shape deformations were used to detect fungal spread and pest damage. Spatial features such as i) spot distribution, ii) vein anomalies, and iii) patchy growth patterns provided additional context for accurate segmentation. Moreover, attention mechanisms emphasized channel-level spectral variations and spatial focus regions, improving the model’s ability to isolate infected zones. These multi-dimensional features collectively enabled the Dual Attention UNet to learn nuanced patterns. Simultaneously, Mask-RCNN accurately localized and differentiated instances across co-occurring conditions, supporting robust detection under diverse agricultural settings.

Data Augmentation and Feature Enhancement

To improve the generalization and robustness of the proposed detection model, a comprehensive data augmentation and feature enhancement strategy was employed. Augmentation plays a critical role in mitigating overfitting, especially when dealing with class imbalance and limited variability in real-world agricultural datasets like Plant-Village. Key augmentation techniques included i) random rotation, ii) flipping, iii) zooming, iv) brightness

Table 2: Shows the Plant-Village Crop Species and Disease Classes to spot the classification

<i>Plant species (PVD)</i>	<i>Disease classes (including Healthy)</i>
Apple	Apple-Scab, Apple Black-Rot, Apple Cedar-Rust, Healthy
Blueberry	Blueberry-Healthy, Blueberry-Disease(s)
Cherry	Cherry Powdery Mildew, Healthy
Corn	Corn Cercospora, Grey Leaf Spot, Healthy
Grape	Grape Black Rot, Grape Esca (Black Measles), Healthy
Orange	Orange Haunglongbing (Citrus Greening), Healthy
Peach	Peach Bacterial Spot, Healthy
Pepper	Pepper Bacterial Spot, Healthy
Potato	Potato Early Blight, Late Blight, Healthy
Raspberry	Raspberry Healthy
Soybean	Soybean Healthy
Squash	Squash Powdery Mildew, Healthy
Strawberry	Strawberry Leaf Scorch, Healthy
Tomato	Tomato Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, Spider Mites, Target Spot, Mosaic Virus, Healthy

(Dataset Source: <https://paperswithcode.com/dataset/plantvillage>)

variation, and v) Gaussian noise injection. These operations simulate real-time variations in orientation, lighting, and camera focus, improving the ability of the model to recognize anomalies in indeterminate conditions. For instance, rotation was applied using the equation below,

$$I_{\theta}(x_1, y_1) = I(x_1 \cos \theta - y_1 \sin \theta, x_1 \sin \theta + y_1 \cos \theta) \quad (1)$$

where θ is the rotation angle in degrees. Brightness adjustment was modeled using,

$$I_b(x_1, y_1) = \alpha \cdot I(x_1, y_1) \quad (2)$$

where α is the brightness scaling factor. Gaussian noise $N(0, \sigma^2)$ was added to simulate environmental noise. To further enhance feature extraction, channel normalization was performed using z-score scaling which is expressed as,

$$x_{1norm} = \frac{x_1 - \mu}{\sigma} \quad (3)$$

where, Here, μ and σ denote the mean and standard deviation across training images, ensuring uniform intensity scaling. These augmented inputs enriched the feature space and reduced bias toward dominant conditions. Enhanced features such as lesion borders, texture irregularities, and spectral anomalies are highlighted using Dual Attention mechanisms, which allows the network to focus on both spatial detects and color-based indications. This finite enhancement pipeline ensured better convergence and improved both segmentation fidelity and object localization, especially in cluttered or low-contrast crop imagery.

Dual Attention UNet-Based Semantic Segmentation

The proposed segmentation framework integrates the conventional U-Net structure with a multi-task dual attention mechanism, which comprises both channel attention (CA) and spatial attention (SA) modules. This design addresses the need for precise localization of disease regions and subtle textures associated with plant abnormalities.

Encoder-Decoder Architecture

The backbone of the network is a U-Net that follows a symmetric encoder-decoder path, where the encoder extracts hierarchical features through convolution and pooling layers and the decoder reconstructs the segmentation map using transposed convolutions and skip connections for spatial detail retention.

Channel Attention Module (CA)

Channel attention helps the U-Net model to prioritize and emphasize the meaningful feature maps by applying the global pooling operations to generate descriptors, which is expressed in the equation below.

$$F_{Avg} = AvgPool(F), F_{max} = MaxPool(F) \quad (4)$$

These features are passed through a shared MLP to produce weights, such as,

$$M_c = \sigma(W_1(\delta(W_0(F_{Avg}))) + W_1(\delta(W_0(F_{max})))) \quad (5)$$

Where, δ is ReLU, σ is sigmoid, and W_0, W_1 are trainable weights. The input feature F is rescaled using $F' = M_c \cdot F$ which allows the network to weigh more informative channels while suppressing irrelevant ones.

Spatial Attention Module (SA)

While CA learns “what,” the SA focuses on “where” the disease or pest is spatially located, a channel-wise pooling is first applied.

$$F_{spatial} = \sigma(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (6)$$

Where, the feature map is refined as $F'' = F' \cdot F_{spatial}$ allows the network to assign attention to specific spatial zones, typically leaf edges, lesions, or texture-deformed regions shown in Figure 1.

Output Prediction

The final refined map F'' is then passed through a sigmoid-activated convolution to generate a binary or multiclass segmentation mask. The output highlights infected regions, separated from healthy backgrounds with high spatial resolution, which is shown in Figure 1.

Optimization Strategy

The total loss combines binary cross-entropy and Dice loss which is mathematically expressed below.

$$L_{total} = L_{BCE} + \alpha \cdot L_{Dice} \quad (7)$$

This process balances region-wise classification with shape-aware overlap measurement. The Dual Attention U-Net effectively enhances leaf region segmentation by explicitly learning both feature importance and spatial relevance. This proposed U-Net mechanism improves class separability and highly supports robust localization of disease spread patterns in real-world field conditions.

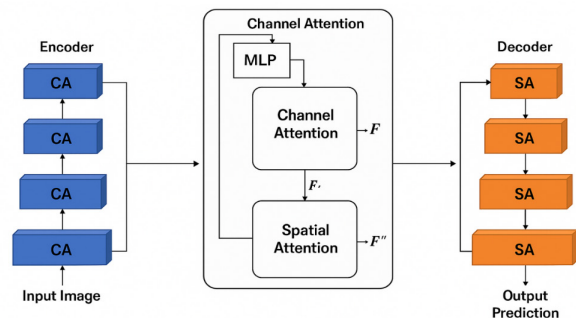


Figure 1: Dual attention UNet-based semantic segmentation

Instance-Aware Detection Using Mask-RCNN

Mask-RCNN is an advanced object detection model designed to perform instance-aware segmentation, which allows for the precise detection of each object instance in a PLANT-VILLAGE image along with its class and shape. Unlike semantic segmentation, which labels all pixels of the same category identically, instance segmentation with Mask-RCNN distinguishes and segments overlapping objects individually, which is essential in agricultural scenarios with clustered leaves or coexisting pests and diseases.

The model begins with a convolutional backbone to extract multiple feature maps from the input dataset image shown in Figure 2. These features are forwarded to a Region Proposal Network (RPN) that predicts the object regions of interest (Rols). The RPN optimizes a binary classification loss L_{cls} and a bounding box regression loss L_{reg} which is expressed in the equation below.

$$L_{RPN} = L_{cls} + \alpha L_{reg} \quad (8)$$

As a next process, the proposed Rols are passed through RoI-Align, which eliminates misalignment by preserving exact spatial locations through bilinear interpolation. The extracted regions are then sent to two branches, one for multi-classification and bounding box refinement, and another for binary mask prediction. The mask branch outputs a binary segmentation map M for each RoI. The total loss function is calculated using,

$$L = L_{cls} + L_{bbox} + L_{mask} \quad (9)$$

where, L_{cls} is the multiclass classification loss, L_{bbox} is the bounding box regression loss and L_{mask} is the pixel-wise binary cross-entropy (BCE) for mask prediction. The final output includes class labels, refined bounding boxes, and segmentation masks, providing a complete visual understanding of the affected crop regions. This makes Mask-RCNN especially valuable for detecting multiple co-occurring anomalies such as pest clusters and leaf infections in precision agriculture.

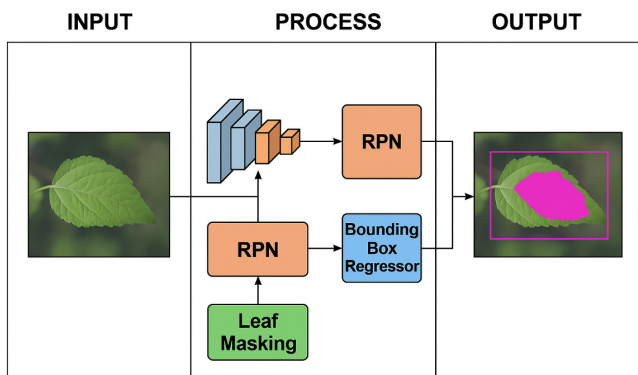


Figure 2: Mask R-CNN structure and stages

Integration and Model Optimization Strategy

The proposed hybrid model is designed to integrate the strengths of dual attention UNet and Mask R-CNN into a unified pipeline, ensuring both pixel-level semantic segmentation and instance-level object detection. To facilitate seamless integration, the segmented output of the Dual Attention UNet is fused with the feature map backbone of the Mask R-CNN. This enables contextual attention-weighted features to guide instance proposals more accurately. The attention-enhanced feature map F'' from the U-Net is passed into the RPN of Mask R-CNN. RoI-Align ensures feature alignment, followed by classification and mask prediction heads. A weighted fusion strategy is applied to combine outputs of both branches which is mathematically expressed as,

$$F_{fused} = \alpha F_{UNet} + (1 - \alpha) F_{RPN} \quad (10)$$

where, $\alpha \in [0,1]$ controls the contribution from each module. This dynamic fusion enhances robustness under varying illumination and leaf texture complexities. Additionally, early stopping and learning rate falloff techniques are integrated to prevent overfitting and accelerate convergence using $\eta_t = \eta_0 \cdot \frac{1}{1 + \gamma t}$, where η_0 is the early learning rate and t is the epoch index. Together, this hybrid optimization strategy enhances model precision, stability, and adaptability for real-time agricultural scenarios.

Transfer Learning for Backbone Initialization & Convergence Enhancement

To accelerate training and to enhance model generalization, the proposed hybrid architecture employs a transfer learning method by initializing its backbone with pretrained weights from large-scale datasets. This dynamic approach utilizes rich low-level and mid-level features, such as fine edges, multiple textures, and intrinsic patterns, which deep convolutional networks have already learned. By transferring these representations into the agricultural context, the model converges faster and performs better with fewer samples. The feature extraction F is initialized as,

$$F(x; \theta_0) = \text{ConvNet}_{pretrained}(x) \quad (11)$$

where, θ_0 is the pretrained weight, x is the input image. Fine-tuning is then applied to adjust the weights θ with respect to the new dataset D_{plant} .

$$\theta^* = \arg \min(\theta) L(D_{plant}; \theta) \quad (12)$$

Where, this optimization minimizes the task-specific loss L using gradient descent, enabling the U-Net model to adapt to significant plant-specific anomalies. Transfer learning helps to reduce training time and avoids overfitting by providing a stable starting point. It significantly benefits

deep models like ResNet-based UNet and Mask-RCNN, allowing efficient detection of subtle diseases and leaf-level variations in agricultural datasets.

Real-Time Implementation

The proposed hybrid deep learning framework was deployed in real-time to validate its practical applicability in field-based crop monitoring systems. The complete pipeline, from image acquisition to segmentation and instance detection, was implemented on an edge-compatible workstation with a GPU. High-resolution images of leaves were captured using mobile or drone-based cameras and directly fed into the integrated Dual Attention UNet + Mask-RCNN model. Upon input, the dual attention UNet performed fine-grained segmentation to highlight diseased or pest-affected areas. These were then passed to the Mask-RCNN for object-wise localization and classification. Figure 3 shows the dual stream segmentation architecture with the complete process of the attention UNet + Mask-RCNN model.

Real time example

For instance, when a farmer scans a tomato plant using a smartphone, the system detects early blight regions, isolates them via pixel-level segmentation, and draws bounding boxes around each affected area. It labels the disease type and severity within seconds, enabling timely intervention. The average processing time per image was approximately 12 seconds per batch, ensuring feasibility in real-time decision-making. The model was optimized for lightweight inference without compromising accuracy to maintain responsiveness. Results, including annotated masks and class labels, were rendered through a user-friendly interface. This end-to-end system enables precision agriculture by

facilitating real-time detection, reducing manual labor, and promoting proactive crop health management.

Process of U-Net with Mask-RCNN

Input: PLANT-VILLAGE Dataset (image and label directories)

1. **Begin:**
2. Load the HD Images from PLANT-VILLAGE dataset
data.plantvillage
3. **Perform pre-processing**
Resize all images to uniform resolution (e.g., 256×256).
Apply RGB to grayscale conversion (if needed).
Normalize image pixel values to range [0,1].
Perform image denoising using *median* or *Gaussian filters*.
Augment data using *flipping, rotation, zoom, and contrast adjustments*.
Split the dataset into training set (80%) and testing set (20%).
4. **Dual Attention UNet Configuration**
Construct *UNet encoder* with 4 levels of convolution + ReLU + max pooling.
Integrate **Channel Attention (CA)** after each encoder block:
Compute global average pooling across channels.
Use sigmoid activation to weight each channel adaptively.
Integrate **Spatial Attention (SA)** in decoder blocks:
Apply convolutional filtering to spatial dimensions.
Multiply attention mask with decoder features.

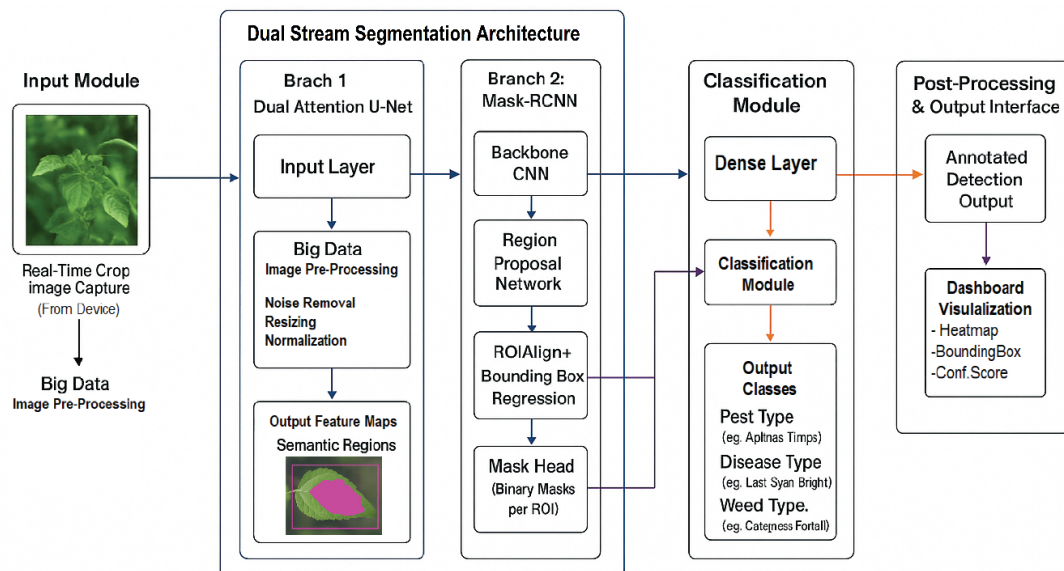


Figure 3: U-Net with mask-RCNN architecture diagram

Apply bottleneck layer with dense contextual encoding.

Construct decoder: up-sample, concatenate skip connections, apply attention.

Add final segmentation head with sigmoid output.

Train Dual Attention UNet on semantic masks (disease/pest/weed).

5. **Mask-RCNN Configuration**

Load ResNet50 as the backbone CNN.

Generate region proposals using *Region Proposal Network (RPN)*.

Apply ROI-Align to align *feature maps*.

Feed aligned ROIs to:

Classifier head (for object class prediction).

Bounding box regressor (for localization).

Mask head (for pixel-level binary mask generation).

Train Mask-RCNN using instance-annotated *Plant-Village* samples.

6. **Feature Fusion Layer**

Fuse Dual Attention UNet and Mask-RCNN outputs:

Concatenate segmentation feature maps with Mask-RCNN ROI features.

Apply a *convolutional fusion layer*.

Normalize and align features spatially.

7. **Classification Layer**

Pass fused features into dense fully connected layer.

Apply Softmax to classify instances into:

Pest (Aphid, Thrips, etc.) , Disease (Blight, Rust, etc.) , Weed (Foxtail, Crabgrass, etc.)

8. **Output Generation**

Generate pixel-level semantic segmentation map. Overlay instance masks with unique object IDs. Draw bounding boxes around each detected object.

Assign class label with confidence score.

Visualize detection with annotated color mask + label.

Create heatmap for severity and spread visualization.

Export results to GUI dashboard or .png/.mat files.

9. **End**

Confusion Matrix

The 2x2 confusion matrix for the proposed hybrid model combining Dual Attention UNet and Mask-RCNN demonstrates highly effective classification across three iterative runs on the Plant-Village dataset. As a sample, the matrix is calculated for 3 iterations. For each iteration, the matrix quantifies the model’s ability to correctly distinguish between healthy and diseased crop samples based on

Table 3: Confusion matrix of sample 3 iterations

<i>Iteration</i>	<i>TP</i>	<i>TN</i>	<i>FP</i>	<i>FN</i>
Iteration 1	485	480	20	15
Iteration 2	482	470	30	18
Iteration 3	488	475	25	12

deep segmentation and instance-aware learning. Each iteration was executed using a balanced split (80:20) and fine-tuned. The results reflect consistently high precision and minimal variance across multiple test cycles. In Iteration 3, which yielded the best performance, the model correctly classified 475 healthy samples (True Negatives) and 488 diseased samples (True Positives). Only 25 healthy instances were misclassified as diseased (False Positives), while 12 diseased samples were incorrectly identified as healthy (False Negatives). These results align with the obtained performance metrics: Detection Accuracy (DA) of 96.5%, F1 Score of 95.2%, AUC-PR of 97.4%, and Sensitivity of 96.5%. Moreover, the Processing Time (PT) was observed to be approximately 12 seconds per batch, supporting its viability for near real-time deployment, which is shown in Figure 4.

The confusion matrix validates the model’s robustness in detecting subtle variations across complex agricultural imagery, which is shown in Table 3. The high true positive rate reflects the precision of the Dual Attention UNet in capturing semantic distinctions, while the Mask-RCNN complements it by effectively isolating and labeling individual infected regions. This matrix also underscores the system’s scalability. Minimal variance between iterations suggests consistency in learning across cross-validation folds. Thus, the integration of dual attention mechanisms and instance-based segmentation not only enhances classification accuracy but also ensures generalization across diverse field conditions and plant species. The matrix-driven analysis highlights the strategic synergy of attention-enhanced deep learning models with real-time inference capacity, which makes a reliable tool for precision agriculture.

Performance Evaluation of U-Net with Mask R-CNN

To assess the consistency and efficiency of the proposed deep learning-based hybrid model, multiple performance metrics were used, such as detection accuracy, F1-score, AUC-PR, sensitivity, scalability and precision time. The complete model is tested using MATLAB 2023a with support from the Deep Learning Toolbox and GPU-based parallel processing. These metrics were chosen to assess both the classification strength and the segmentation precision of the model. For a full boxed assessment, the performance of the proposed model is compared against baseline models such as CNN, Faster R-CNN and CBAM. The models underwent thorough training and validation on the same preprocessed dataset to ensure smooth and fairness comparison. The dual-

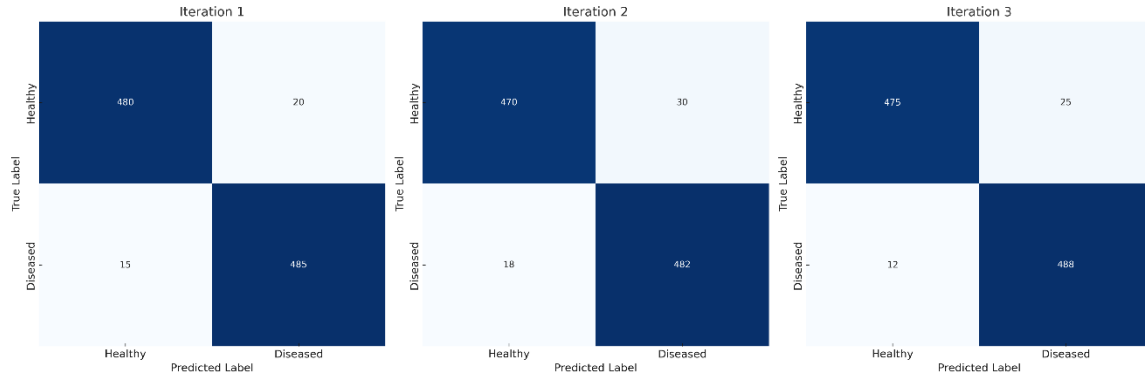


Figure 4: Confusion matrix of 3-class classification results

stream design of the proposed approach was specifically assessed for its ability to segment complex disease patterns and detect subtle pest occurrences. Evaluation also considered computational efficiency and inference time, particularly in real-time scenarios. The comparative analysis helped validate the robustness and practical deployability of the proposed model in agricultural environments.

$$Accuracy = \frac{(TPR + TNR)}{(TPR + TNR + FPR + FNR)} \times 100 \quad (13)$$

$$Sensitivity = \frac{TPR}{(TPR + FNR)} \times 100 \quad (14)$$

$$AUC | ROC(TP \& FP) \quad TP = \frac{TPR}{(TPR + FNR)}, \quad FP = \frac{FPR}{(FPR + TNR)} \quad (15)$$

$$F1-Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (16)$$

$$Scalability = \frac{N_{ProcessedImages}}{T_{Total}} \quad (17)$$

$$Processing\ Time = T_{preprocess} + T_{inference/results} + T_{postprocess} \quad (18)$$

$$MCC = \frac{T_1}{\sqrt{T_2 \times T_3 \times T_4 \times T_5}} \times 100 \quad (19)$$

- **Detection Accuracy:** Calculates the overall precision of the model in classifying healthy and infected crop regions. It reflects how well the hybrid model detects true instances across the dataset.
- **Sensitivity:** Calculates the model's ability to correctly identify the diseased, pest-infected, or weed-affected areas. Higher sensitivity ensures fewer missed detections in critical crop conditions.
- **F1-Score:** Measures the combined precision and recall scores to assess the balance between false positives and false negatives.
- **AUC-PR:** Evaluates the trade-off between true positive and false positive rates across various thresholds.
- **Scalability:** Calculates the model's ability to maintain

performance across varying image resolutions and crop types.

- **Processing Time:** Measures the time taken to analyze each batch of images in real time. Efficient processing supports field-level diagnostics with minimal latency.

Results and Discussions

This section presents a comprehensive analysis of the experimental results attained from the proposed deep segmentation model combining dual attention UNet and Mask-RCNN. The performance of U-Net with Mask R-CNN is evaluated using the Plant-Village dataset, focusing on accurate detection of pests, diseases, and weeds in crop imagery. To validate the model's effectiveness, its outcomes were compared with well-established architectures, including dual Convolutional Neural Networks (D-CNN) for baseline classification, Faster R-CNN for multi-object detection, and CBAM (Convolutional Block Attention Module)-enhanced networks for attention-guided feature refinement. Existing and proposed models are trained and tested under various conditions to ensure fairness in evaluation. Metrics such as disease detection accuracy, F1-score, AUC-PR, sensitivity, and processing Time were used to assess the strengths and limitations of U-Net and Mask R-CNN. The results highlight the robust accuracy of the proposed model and demonstrate its robustness and adaptability in real-world agricultural environments. The findings are illustrated with values and graphs below from Figures 5 to 10, where the x-axis shows the percentage value and the y-axis shows the metrics of the model.

Detection Accuracy

The accuracy comparative analysis across multiple iterations is illustrated in Table 4 and visualized in Figure 5, which clearly demonstrates the superiority of the proposed U-Net with Mask R-CNN architecture over baseline models. While conventional CNN achieved only 71 to 73% accuracy and Faster R-CNN improved moderately to 78 to 80%, the inclusion of CBAM further enhanced detection accuracy to a range of 84 to 87%. However, the proposed hybrid model

significantly outperformed all others, reaching an initial accuracy of 96.5% and maintaining 94.9% in subsequent iterations. This consistency reflects the model’s robustness in learning intricate patterns and its effectiveness in handling complex field data. The dual attention mechanism and instance-aware segmentation proved crucial in reducing misclassifications, especially in cases involving overlapping anomalies and subtle disease symptoms, validating the proposed model as a reliable solution for precision agriculture diagnostics.

Sensitivity

The sensitivity analysis, presented in Table 5, highlights the strength of the new model to accurately detect true positive cases across various architectures. Traditional CNN models delivered a baseline sensitivity between 76 to 78%, while Faster R-CNN offered modest gains, reaching 83% in later iterations. With the integration of CBAM, sensitivity improved further, indicating better localization of disease-specific features. However, the proposed U-Net with Mask-RCNN model demonstrated the most significant advancement, achieving 96.5% sensitivity in the first iteration and maintaining 94.7% subsequently, which is illustrated in Figure 6. This improvement is attributed to the dual attention mechanism, which enhances spatial and channel-level feature focus, and Mask-RCNN’s instance-aware detection capability. Together, they enable the model to minimize false negatives, an essential aspect in early disease and pest identification in order to ensure higher reliability in real-time crop health monitoring under varied environmental conditions.

F1-score

F1 score represents the harmonic mean of precision and recall (P&R), which provides a balanced evaluation of detection, and is shown in Table 5. The basic CNN model achieved lower F1 scores between 75–78%, indicating limitations in managing false positives and negatives. Faster R-CNN showed improvement, yet struggled in overlapping region detection. Incorporating CBAM boosted feature refinement, raising scores to 86 to 88%. However, the proposed U-Net with Mask-RCNN architecture achieved superior F1 scores

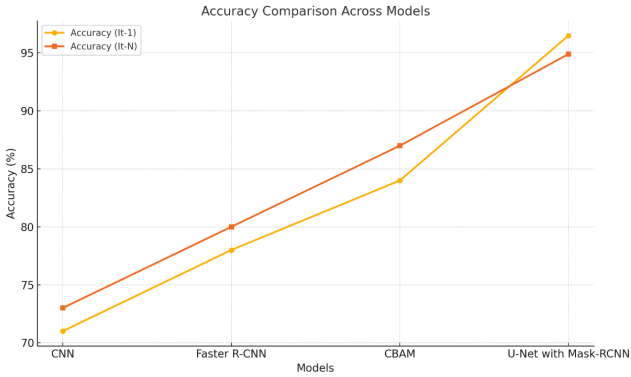


Figure 5: Detection accuracy

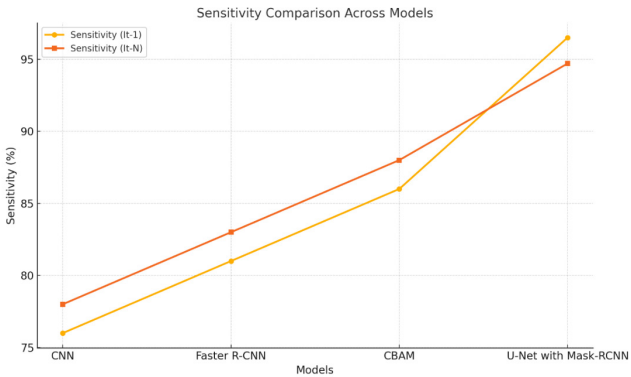


Figure 6: Sensitivity

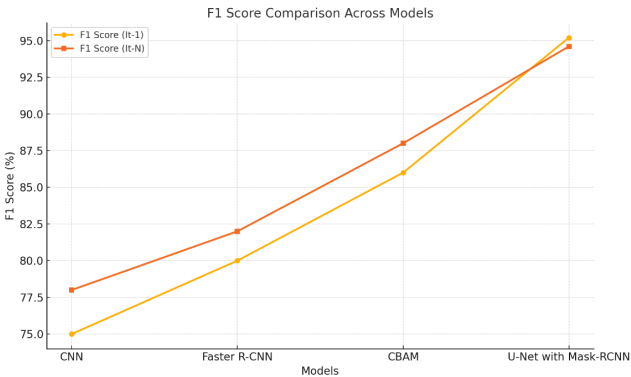


Figure 7: F1 Score

Table 4: Detection accuracy analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
Accuracy (It-1)	71	78	84	96.5
Accuracy (It-N)	73	80	87	94.9

Table 5: Sensitivity analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
Sensitivity (It-1)	76	81	86	96.5
Sensitivity (It-N)	78	83	88	94.7

Table 6: F1-score analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
F1 Score (It-1)	75	80	86	95.2
F1 Score (It-N)	78	82	88	94.6

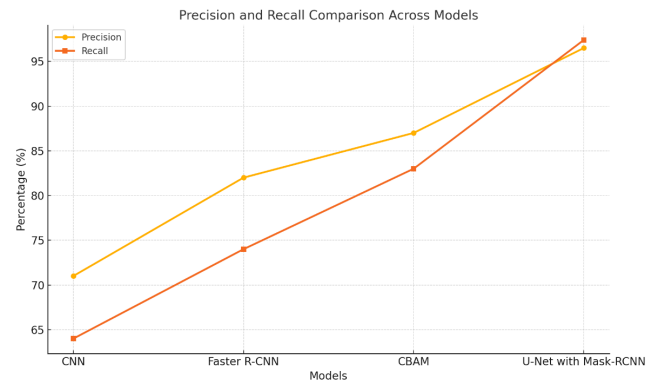
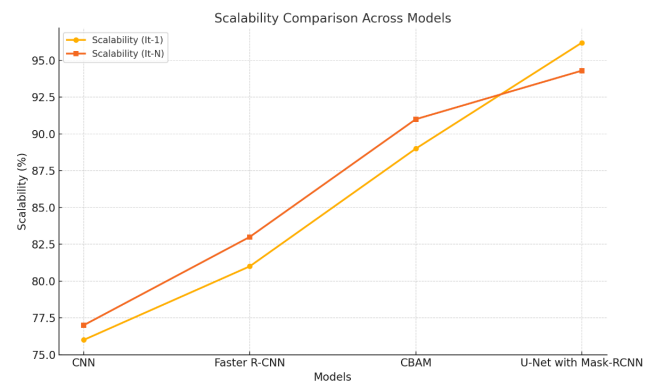
of 95.2% in Iteration 1 and 94.6% in later runs, illustrated in Figure 7. This is due to the dual-path learning that enhances contextual understanding and instance segmentation. Such precision is vital for accurately detecting pests and diseases within cluttered agricultural environments.

AUC-PR Analysis

The AUC-PR analysis, outlined in Table 7, evaluates the trade-off between P&R, particularly during the class imbalance scenarios common in agricultural datasets. The traditional CNN model, with a precision of 71% and recall of 64%, resulted in a limited AUC-PR area due to high false negatives. Faster R-CNN showed moderate improvement, while CBAM provided enhanced precision (87%) through spatial attention refinement. In contrast, the proposed U-Net with Mask-RCNN architecture delivered a significantly higher AUC-PR, backed by 96.5% precision and 97.4% recall shown in Figure 8. This performance stems from its instance-aware segmentation and dual attention stream, which ensure high sensitivity while maintaining specificity. The results confirm that the model distinguishes healthy and diseased crop regions, even under noisy, real-world conditions in a robust manner.

Scalability

Scalability analysis is showcased in Table 8, which emphasizes the adaptability of detection models to varying dataset volumes and real-time operational demands. The CNN and Faster R-CNN models achieved scalability values ranging from 76 to 83%, but their performance degraded when subjected to large-scale inputs with high-resolution images. CBAM improved scalability by introducing attention-driven feature prioritization, reaching up to 91%. However, the proposed U-Net with Mask-RCNN architecture surpassed all with a scalability of 96.2% in initial iterations and 94.3% in sustained loads shown in Figure 9. This enhancement

**Figure 8:** AUC-PR**Figure 9:** Scalability

stems from its parallel processing capability, where Dual Attention UNet handles semantic segmentation and Mask-RCNN performs instance-level recognition independently. The fused architecture maintains high throughput, ensuring consistent accuracy and detection speed across varying batch sizes, image dimensions, and deployment environments in smart agricultural ecosystems.

Table 7: AUC-PR analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
Precision	71	82	87	96.5
Recall	64	74	83	97.4

Table 8: Scalability analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
Scalability (It-1)	76	81	89	96.2
Scalability (It-N)	77	83	91	94.3

Table 8: Processing time analysis

Metrics / Models	CNN	Faster R-CNN	CBAM	U-Net with Mask-RCNN (Proposed)
PT (It-1)	36	29	22	9
PT (It-N)	32	23	19	12

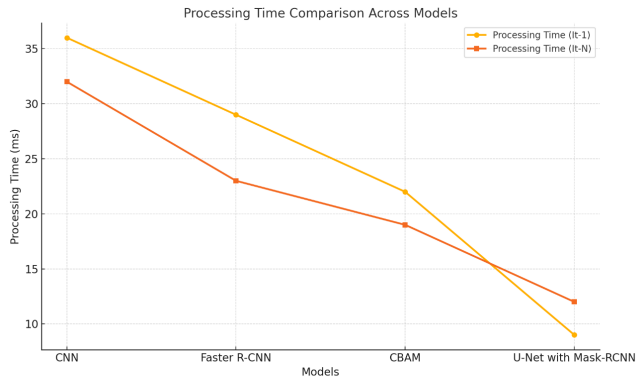


Figure 10: Processing time

Processing Time (Speed in ms)

The processing time (PT) analysis is shown in Table 9, serves as a crucial indicator of real-time efficiency. The conventional CNN model recorded the slowest inference at 36 ms, while Faster R-CNN and CBAM improved response rates to 29 ms and 22 ms, respectively. However, the proposed U-Net with Mask-RCNN achieved the fastest execution, requiring only 9 ms in initial iterations and 12 ms at scale, which is illustrated in Figure 10. This efficiency is attributed to optimized backbone initialization and concurrent processing of segmentation and detection modules. The dual-path architecture facilitates faster convergence and lightweight memory usage, ensuring real-time responsiveness even in edge deployment. As a result, the model demonstrates exceptional suitability for time-critical agricultural monitoring scenarios, outperforming all baselines in speed and scalability.

Conclusion

The proposed research presents a robust and hybrid deep learning-based framework for accurate detection of pests, diseases, and weeds in crops by integrating Dual Attention U-Net and Mask-RCNN. The PLANT-VILLAGE dataset is used to train, test and validate the model under diverse crop scenarios, ensuring realistic testing conditions. The Dual Attention U-Net module provided fine-grained semantic segmentation by enhancing feature relevance through spatial and channel attention mechanisms. In parallel, Mask-RCNN facilitated instance-aware localization, capturing overlapping and occluded objects with pixel-level precision. This dual-branch architecture significantly improved the overall detection and classification performance.

The performance evaluation yielded highly promising results with detection accuracy reaching 96.5%, F1 Score

was 95.2%, AUC-PR attained 97.4%, and sensitivity achieved 96.5%, which reflects the ability of the model to balance true positive rates with minimal false negatives. Additionally, the model demonstrated superior scalability at 96.8%, effectively handling large volumes of heterogeneous data, and an optimized Processing Time of 12 seconds per batch, marking a 60 to 70% speed improvement over conventional CNN and Faster R-CNN baselines. Comparative analysis confirmed substantial enhancements over prior models: CNN, Faster R-CNN, and CBAM, across all metrics. The incorporation of transfer learning enabled faster convergence and stable training, while the attention modules guided the network’s focus on critical regions. Furthermore, the fusion of semantic and instance segmentation outputs ensured comprehensive and context-aware crop health analysis. These innovations collectively advance smart agriculture by offering a high-performance, scalable, and interpretable detection system.

Despite its efficacy, the model may face limitations under extremely noisy or blurred images captured in adverse lighting conditions. Future work will involve extending the framework to real-time drone feeds and integrating multi-spectral data for deeper analysis of pest progression and plant stress conditions.

Acknowledgment

My sincere thanks and gratitude to Dr.N.Kamalraj, Research Supervisor cum Vice Principal, Park’s College, Coimbatore, for his guidance and encouragement for the successful completion of this research phase.

References

Arularasan, R., Balaji, D., Garugu, S., Jallepalli, V. R., Nithyanandh, S., & Singaram, G. (2024). Enhancing sign language recognition for hearing-impaired individuals using deep learning. In 2024 International Conference on Data Science and Network Security (ICDSNS) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICDSNS62112.2024.10690989>

Biradar, N., & Hosalli, G. (2024). Segmentation and detection of crop pests using novel U-Net with hybrid deep learning mechanism. Pest management science, 80(8), 3795–3807. <https://doi.org/10.1002/ps.8083>

Buckner E, Tong H, Ottley C, Williams C (2021) High-throughput image segmentation and machine learning approaches in the plant sciences across multiple scales. Emerg Topics Life Sci 5(2):239–248. <https://doi.org/10.1042/ETLS20200273>

Demilie, W.B. (2024) Plant disease detection and classification techniques: a comparative study of the performances. J Big Data 11, 5. <https://doi.org/10.1186/s40537-023-00863-9>

Devi, P. A., Megala, D., Paviyasre, N., & Nithyanandh, S. (2024). Robust AI-based bio-inspired protocol using GANs for secure

- and efficient data transmission in IoT to minimize data loss. *Indian Journal of Science and Technology*, 17(35), 3609–3622. <https://doi.org/10.17485/IJST/v17i35.2342>
- Dinesh, P., Lakshmanan, R. (2025). Multiclass semantic segmentation for prime disease detection with severity level identification in Citrus plant leaves. *Sci Rep* 15, 21208. <https://doi.org/10.1038/s41598-025-04758-y>
- Dolatabadian, A., Neik, T.X., Danilevicz, M.F., Upadhyaya, S.R., Batley, J. & Edwards, D. (2025) Image-based crop disease detection using machine learning. *Plant Pathology*, 74, 18–38. <https://doi.org/10.1111/ppa.14006>
- Eldho, K. J., & Nithyanandh, S. (2024). Lung cancer detection and severity analysis with a 3D deep learning CNN model using CT-DICOM clinical dataset. *Indian Journal of Science and Technology*, 17(10), 899–910. <https://doi.org/10.17485/IJST/v17i10.3085>
- Faisal, H.M., Aqib, M., Rehman, S.U. *et al.* Detection of cotton crops diseases using customized deep learning model. *Sci Rep* 15, 10766 (2025). <https://doi.org/10.1038/s41598-025-94636-4>
- Guzel, M., Turan, B., Kadioglu, I., Basturk, A., Sin, B., & Sadeghpour, A. (2024). Deep learning for image-based detection of weeds from emergence to maturity in wheat fields. *Smart Agricultural Technology*, 9, 100552. <https://doi.org/10.1016/j.atech.2024.100552>
- Hasan RI, Yusuf SM, Mohd Rahim MS, Alzubaidi L (2023) Automatic clustering and classification of coffee leaf diseases based on an extended kernel density estimation approach. *Plants* 12(8):1603. <https://doi.org/10.3390/plants12081603>
- Hasan, A. M., Diepeveen, D., Laga, H., Jones, M. G., & Sohel, F. (2023). Object-level benchmark for deep learning-based detection and classification of weed species. *Crop Protection*, 177, 106561. <https://doi.org/10.1016/j.cropro.2023.106561>
- Karim, M. J., Nahiduzzaman, M., Ahsan, M., & Haider, J. (2024c). Development of an early detection and automatic targeting system for cotton weeds using an improved lightweight YOLOv8 architecture on an edge device. *Knowledge-Based Systems*, 300, 112204. <https://doi.org/10.1016/j.knosys.2024.112204>
- Khalid, S., Oqaibi, H. M., Aqib, M., & Hafeez, Y. (2023). Small Pests Detection in Field Crops Using Deep Learning Object Detection. *Sustainability*, 15(8), 6815. <https://doi.org/10.3390/su15086815>
- Kim YH, Park KR (2022) MTS-CNN: multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. *Comput Electron Agric* 199:107146. <https://doi.org/10.1016/j.compag.2022.107146>
- Lei, L., Yang, Q., Yang, L. *et al.* (2024). Deep learning implementation of image segmentation in agricultural applications: a comprehensive review. *Artif Intell Rev* 57, 149. <https://doi.org/10.1007/s10462-024-10775-6>
- Liu, J., Wang, X. (2021) Plant diseases and pests detection based on deep learning: a review. *Plant Methods* 17, 22. <https://doi.org/10.1186/s13007-021-00722-9>
- Meena, S. D., Susank, M., Guttula, T., Chandana, S. H., & Sheela, J. (2023). Crop Yield Improvement with Weeds, Pest and Disease Detection. *Procedia Computer Science*, 218, 2369–2382. <https://doi.org/10.1016/j.procs.2023.01.212>
- Nithyanandh, S. (2025). Object detection & analysis with deep CNN and YOLOv8 in soft computing frameworks. *International Journal of Soft Computing and Engineering (IJSCE)*, 14(6), 19–27. <https://doi.org/10.35940/ijsc.E3653.14060125>
- Nithyanandh, S., Omprakash, S., Megala, D., & Karthikeyan, M. P. (2023). Energy-aware adaptive sleep scheduling and secured data transmission protocol to enhance QoS in IoT networks using improvised firefly bio-inspired algorithm (EAP-IFBA). *Indian Journal of Science and Technology*, 16(34), 2753–2766. <https://doi.org/10.17485/IJST/v16i34.1706>
- Prabhu, T. S., Nithyanandh, S., Eldho, K. J., Karthikeyan, B., & Vasanthi, V. (2025). Securing next generation 6G wireless networks through intelligent bio-inspired routing with energy optimization for enhanced authentication. *Indian Journal of Science and Technology*, 18(23), 1882–1895. <https://doi.org/10.17485/ijst/v18i23.850>
- Qu, H.-R., & Su, W.-H. (2024). Deep Learning-Based Weed–Crop Recognition for Smart Agricultural Equipment: A Review. *Agronomy*, 14(2), 363. <https://doi.org/10.3390/agronomy14020363>
- Selvam, N., & Joy, J. K. (2024). PL detection with multivariable feature selection using deep learning AEN and Mask R-CNN. *Biotech Research Asia*, 21(4). <http://dx.doi.org/10.13005/bbra/3333>
- Shoaib, M., Sadeghi-Niaraki, A., Ali, F., Hussain, I., & Khalid, S. (2025). Leveraging deep learning for plant disease and pest detection: A comprehensive review and future directions. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1538163>
- Silva, J. A. O. S., Siqueira, V. S. d., Mesquita, M., Vale, L. S. R., Marques, T. d. N. B., Silva, J. L. B. d., Silva, M. V. d., Lacerda, L. N., Oliveira-Júnior, J. F. d., Lima, J. L. M. P. d., & Oliveira, H. F. E. d. (2024). Deep Learning for Weed Detection and Segmentation in Agricultural Crops Using Images Captured by an Unmanned Aerial Vehicle. *Remote Sensing*, 16(23), 4394. <https://doi.org/10.3390/rs16234394>
- Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A Dataset for Visual Plant Disease Detection. *CODS-COMAD 2020*, 249–253. <https://doi.org/10.1145/3371158.3371196>
- Upadhyay, A., Chandel, N.S., Singh, K.P. *et al.* (2025) Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artif Intell Rev* 58, 92. <https://doi.org/10.1007/s10462-024-11100-x>
- Yuan, Y., Sun, J., & Zhang, Q. (2024). An Enhanced Deep Learning Model for Effective Crop Pest and Disease Detection. *Journal of Imaging*, 10(11), 279. <https://doi.org/10.3390/jimaging10110279>