ISSN: 2093-5374 / E-ISSN: 2093-5382

A Context-Aware IoT and Edge Computing Framework for Wireless Plant Disease Diagnosis Using Compressed MaskRCNN and ResNet-50

K. Sowmiya[1*], and V.P. Anitha[2]

[1*]Department of Computer Science and Engineering, SRM Institute of Technology and Science, Chengalpattu, Chennai, India. sk6823@srmist.edu.in, https://orcid.org/0000-0003-3106-8860

[2]Assistant Professor, Department of AI&ML, Vels Institute of Science, Technology & Advanced Studies, Pallavaram, Chennai, India. vphoneytha@gmail.com, https://orcid.org/0009-0008-0563-5423

Received: March 30, 2025; Revised: May 13, 2025; Accepted: June 10, 2025; Published: June 30, 2025

Abstract

Plant disease diagnosis remains a challenging problem in contemporary agriculture, as it relies on physical examination, which is labor-intensive, time-consuming, and often inaccurate at industrial or farmland sites. Timely, precise, and maskable diagnostic activity is increasingly needed to support the improvement of precision agriculture and food security worldwide. To address the need for fast, scalable, and trustworthy diagnostics, this paper introduces a context-aware IoT and Edge Computing solution based on a pruned and knowledge-distilled compressed Mask R-CNN model with a ResNet-50 backbone for real-time plant disease diagnosis. Pruning and knowledge distillation enable the system to be efficiently deployed on low-end edge devices, such as the Raspberry Pi and Jetson Nano. Contextual environmental data, such as temperature, humidity, and moisture, are combined with visual input to improve diagnostic performance under different field conditions. Low-power, low-latency wireless communication is facilitated through MQTT and dynamic frequency transmission based on detection events. The model was trained and validated on a benchmark dataset of diseases in tomato and sugarcane leaves, achieving 91.6% classification accuracy and an F1-score of 90.7%, with only a 2.5% accuracy loss compared to the uncompressed model. The inference latency was reduced to 220 ms in edge devices, with a 38% decrease in power consumption, all for eco-friendly operation. These findings validate the applicability of deep learning models in monitoring plant health in real, low-connectivity scenarios. The proposed solution facilitates early intervention, prevents pesticide misuse, and promotes a data-driven vision for smart agriculture.

Keywords: Plant Disease Diagnosis, IoT, Edge Computing, Mask R-CNN, ResNet-50, Agriculture, Classification, Deep Learning.

1 Introduction

The agricultural industry accounted for half of India's employment and 19.9% of the country's gross domestic product in the fiscal year 2020–21. The most recent technical developments must support the high-yield agriculture sector. Insects and plant diseases cause significant crop losses in the agricultural

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 16, number: 2 (June), pp. 707-720. DOI: *10.58346/JOWUA.2025.I2.043*

*Corresponding author: Department of Computer Science and Engineering, SRM Institute of Technology and Science, Chengalpattu, Chennai, India.

industry. By 2050, the world's population is projected to reach 9.2 billion, necessitating a 69.6% increase in food production (Friedrich, 2015). Wind, adverse weather conditions, drought, fungus, bacteria, viruses, and other pathogens cause significant damage to food crops. Plant diseases are the leading cause of crop losses globally, accounting for 70-80% of all losses (Wang et al., 2025). Plant diseases cause a significant amount of crop loss, which needs to be addressed. To ensure healthy and productive food crops, early detection of plant diseases enables farmers to develop effective defense mechanisms against them (Shi et al., 2024). Disease identification and eradication will further increase productivity. The Food and Agriculture Organization estimates that at least 40% of global agricultural production is lost annually due to plant diseases, resulting in significant economic losses, food shortages, and disruptions to agrarian trade (Silva & Almeida, 2024). These effects are particularly devastating in the developing world, where working farmers often lack access to early diagnosis, specialized agronomic advice, and advanced disease management technology. Plant diseases can spread rapidly, devastate entire fields, and undermine food security both locally and nationally in the absence of early detection and effective control (Rehman et al., 2025).

Symptoms of yellow leaves, dark lesions, wilting leaves, and stunted plant growth associated with plant disease often become visually indistinguishable, making diagnosis at the field level difficult without the aid of laboratories (Mohan et al., 2023). Usually, when symptoms become apparent, the disease has progressed to a stage where it is no longer curable. Early, accurate, and timely plant disease identification is therefore vital to facilitate quick response measures, control pesticide abuse, reduce crop loss, and enable sustainable agriculture. In an era of climate unpredictability, which further increases the prevalence and volatility of diseases, it is more important than ever to invest in smart plant health monitoring systems for a more robust agricultural system of the future (Ganesh & Kannadhasan, 2025). Conventionally, plant disease diagnosis relies on visual examination by farmers or agronomists, which is subjective, prone to error, and low-cost but not timely for extensive monitoring (González-Briones et al., 2025). Variability in the expression of the disease under different environmental conditions or growth stages of the plants makes conventional methods challenging, highlighting the necessity for automated, real-time, and context-sensitive solutions.

The emergence of the Internet of Things (IoT) and Edge Computing has transformed precision agriculture with the capability to access data in real-time and make localized decisions (Yemunarane et al., 2024). IoT devices, such as cameras and weather sensors, ingest high-frequency farm field data (Monica Nandini, 2024). Edge computing processes this data at or near the source, removing latency, bandwidth, and reliance on cloud infrastructure significantly. This is particularly valuable in rural farm regions where connectivity is limited. Yet, deep learning models for disease identification from images are computationally expensive and thus not directly implementable on edge devices (Udayakumar et al., 2023). In this work, we propose the use of compressed deep learning models, i.e., a lightweight backbone Mask R-CNN with ResNet-50, optimized for effective plant disease detection in tomato and sugarcane crops under resource-limited conditions (Dhurgadevi et al., 2022). This infrastructure enables wireless, real-time diagnosis, making more efficient and sustainable agriculture possible.

1.1 Contribution of the Research

The key contributions of the paper are as follows:

- The paper suggests a real-time plant disease diagnosis IoT-edge system that gathers and processes visual and environmental information in situ. It minimizes dependence on cloud infrastructure and facilitates quicker, local decision-making. The system is scalable and supports low-latency runtimes in rural agribusiness.

- A deep learning model, compressed with Mask R-CNN and ResNet-50 backbone, is trained for the classification and segmentation of diseases. The model allows accurate detection of diseased tissue in plant leaves. This provides efficient visual interpretation under varied disease conditions.
- Quantization and pruning are applied to minimize the model size and computation cost, enabling it to run on the edge. The compression codes allow real-time inference on low-power devices. This means power efficiency and run feasibility in resource-limited environments.
- Environmental contexts, such as temperature, humidity, and soil moisture, are incorporated into visual information to enable adaptive diagnosis. The context-aware system promotes robustness and relevance under varying field conditions. It facilitates better prediction and intelligent farm management.

The paper is structured as follows: Section 2 discusses the background of IoT and deep learning used in plant disease detection (Bagga & Goyal, 2025). Section 3 discusses the edge-IoT method, which is proposed by utilizing a compressed Mask R-CNN with ResNet-50. This section provides an overview of the dataset and preprocessing methods, model compression and adaptation methods, and a wireless communication protocol. Section 4 discusses the experimental setup, and Section 5 presents the performance results. Section 6 includes limitations and challenges, Section 7 includes future work, and Section 8 contains concluding remarks.

2 Literature Review

The area of plant disease diagnosis has witnessed tremendous progress with the deployment of machine learning and deep learning methods (Zhang et al., 2025). Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) were conventional methods that utilized hand-designed features, such as color, shape, and texture, derived from leaf images. Such methods were not robust in noisy or harsh environments. In parallel, Convolutional Neural Networks (CNNs) have demonstrated superior capabilities in automatically learning hierarchical features. YOLO (You Only Look Once), Faster R-CNN, and Mask R-CNN, top object detection models, have continued to enhance the localization and classification of plant pathogens (Sun & Li, 2024). While the models are precise, they are intensive to compute and are well-suited to run on cloud infrastructure. Parallel to this, IoT-based agricultural monitoring platforms have emerged to automatically collect environmental and plant health data remotely. The platforms typically use sensors to measure temperature, humidity, soil moisture content, and photosynthesis in leaves, and transmit these data to cloud servers for processing. Although suitable for large-scale monitoring, such platforms are characterized by high latency and network dependency, particularly in rural areas where the network is poorly accessible.

To counteract cloud constraints, studies have suggested that Edge Computing can enable local data processing, thereby reducing delays in obtaining actionable insights. Edge AI has been utilized in research work on precision irrigation in agriculture, pest detection, and crop health monitoring using lightweight CNNs on Raspberry Pi and Jetson Nano. The complete integration of segmentation and classification models for real-time edge disease detection remains limited (Wang et al., 2022). Computation overhead reduction has led to model compression methods, such as pruning, quantization, and knowledge distillation, to significantly shrink model size without compromising accuracy. These are essential to deployment in energy-restricted edge environments. High latency, excessive bandwidth usage, and the lack of context-aware adaptation continue to characterize most current solutions. Few systems incorporate segmentation and classification with real-time environmental perception to provide adaptive low-power solutions. This work bridges these gaps with a compact Mask R-CNN using ResNet-50 embedded in an IoT-edge platform for real-time context-aware plant disease diagnosis (Sharma et al., 2020).

3 Proposed Methodology

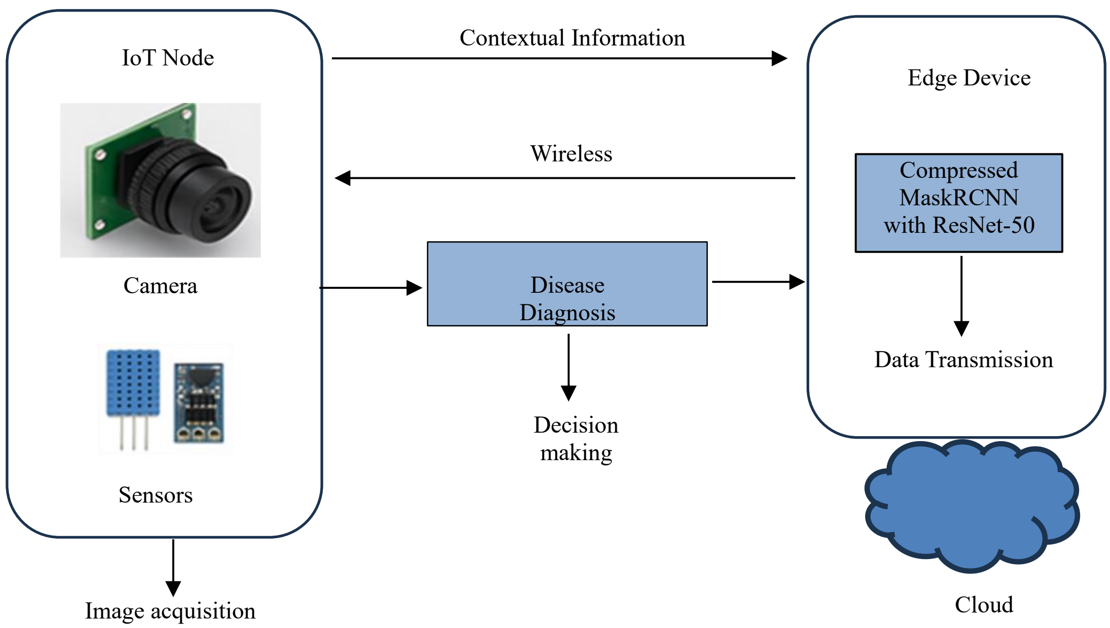

Figure 1: Architecture of the Compressed Mask R-CNN-Based IoT-Edge System for Plant Disease Detection

The architecture for the systems presented in Figure 1 is an overall IoT-edge system designed for context-aware diagnosis of plant diseases, utilizing a compressed Mask R-CNN based on ResNet-50 as the backbone (Mohanty et al., 2016). The main system in this case consists of an IoT node interfaced with a wired camera and various environmental sensors, including humidity, temperature, and soil moisture sensors. These sensors collect visual and contextual information in the field, which constitutes the foundation of smart analysis. The images of leaves are first processed by a compressed Mask R-CNN model, which performs pixel-level segmentation to identify disease-infected areas. Model-based segmentation is optimized for maximum efficiency by utilizing model compression techniques, enabling efficient processing on low-end hardware. Once the infected zones are identified, the segmented output is passed on to an edge device, where the compressed ResNet-50 model is used to determine the precise disease, such as early blight, bacterial wilt, or leaf spot. The employment of environmental context, i.e., geolocation and crop type, enables the model to adjust its inference behavior in real-time, prioritizing humidity-sensitive patterns in wet environments or geolocation-dependent thresholds. This context-dependent adaptation enhances the system's diagnostic accuracy and applicability in different agricultural environments. Upon processing, the final diagnosis and key metadata are wirelessly sent to a cloud server or remote monitor dashboard using lightweight communication protocols. This provides low latency, low bandwidth consumption, and energy conservation.

1.2 Dataset and Preprocessing

The proposed system was trained and tested on a publicly available dataset of plant disease images, specifically for common diseases affecting tomatoes and sugarcane, as shown in Figures 2, 3, and Table 1. It contains high-resolution RGB images of various disease states, including early blight, bacterial wilt, and leaf spots, as well as healthy leaf samples (Li et al., 2025). Each image is labeled with corresponding class labels and, where segmentation tasks are involved, pixel-level masks that delineate the diseased areas. To enhance model generalization and support real-world variations in leaf pose, lighting, and background, aggressive data augmentation strategies were employed. These comprised random horizontal and vertical flipping, rotation (±15°), zooming, and adjustments in brightness/contrast. These extensions simulate field conditions and help prevent overfitting, particularly given the relatively small number of disease-specific image collections.

All images were resized to a uniform input resolution of 512×512 pixels to achieve a balance between computational efficiency and model accuracy. Pixel intensity values were also normalized to the range of 0 to 1 to facilitate training stability. Labeling procedures for the Mask R-CNN model involved creating region proposals and binary masks over each diseased region, which were then manually checked for correctness. The dataset was also divided into 70% training, 15% validation, and 15% test sets with disease category class balance. This preprocessing ensured high-quality input data for training the compressed deep learning model used in the IoT-edge system (Xu & Zhao, 2024).

Table 1: Plant Disease Dataset

Crop	Class	Total Images	Training (70%)	Validation (15%)	Testing (15%)
Tomato	Early Blight	600	420	90	90
Tomato	Late Blight	500	350	75	75
Tomato	Healthy	400	280	60	60
sugarcane	Bacterial Wilt	500	350	75	75
sugarcane	Brown Spot	400	280	60	60
sugarcane	Healthy	400	280	60	60
Total	—	2800	1960	420	420

(a) Early blight **(b) Late blight** **(c) Healthy**

Figure 2: Sample Image of Tomato Leaf Disease

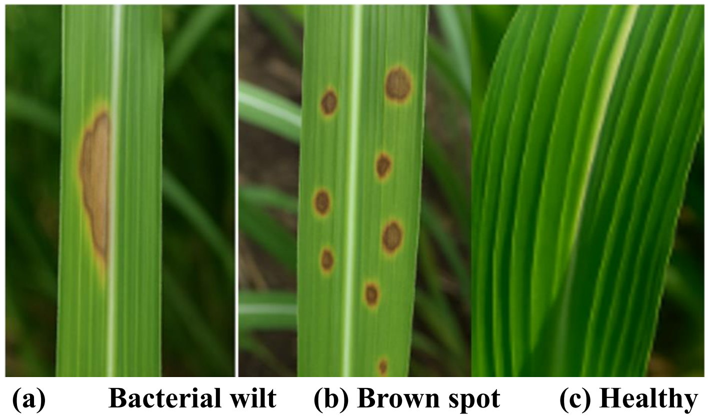

Figure 3: Sample Image of Sugarcane Leaf Disease

1.3 Compressed MaskRCNN with ResNet-50

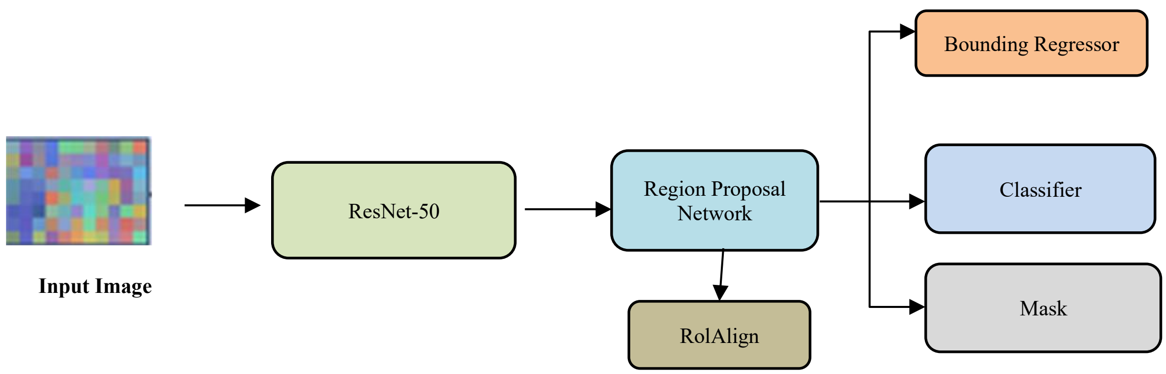

Figure 4: Architecture of the Compressed Mask R-CNN with ResNet-50 Backbone

The core of the proposed disease detection model is built upon the Mask R-CNN architecture with the ResNet-50 backbone presented in Figure 4 (Cheeti et al., 2025). Mask R-CNN is an upgrade to Faster R-CNN, featuring an additional pixel-wise segmentation branch that ensures classification, bounding box regression, and generation of the mask for the diseased area of leaf images are performed simultaneously. The ResNet-50, a 50-layered residual network, serves as the feature extractor. It employs skip connections that address the vanishing gradient problem, facilitating deep feature learning, which is crucial for identifying fine-grained disease patterns between plants, for instance, tomato and sugarcane.

To make this architecture resource-friendly for edge devices, we utilized model compression methods to decrease both memory and computational expenses. Specifically, we applied structured pruning, where we eliminated entire filters and their accompanying feature maps simultaneously, in accordance with the filter with the smallest L1-norm magnitude. Mathematically, for a filter $F_i \in \mathbb{R}^{c \times h \times w}$,its ranking is computed as:

$$\parallel F_i \parallel = \sum_{j=1}^{c} \sum_{k=1}^{h} \sum_{l=1}^{w} |F_i(j,k,l)| \quad (1)$$

The filters with the smallest $\parallel F_i \parallel_1$ values are pruned to yield a sparser, faster network with lower latency.

We also used knowledge distillation, in which a large, precise teacher model $T(x)$ was used to learn a smaller student model $S(x)$ to acquire softened outputs by reducing Kullback–Leibler (KL) divergence:

$$\mathcal{L}_{KD} = \sum_i T_i(x)\,.\log\left(\frac{T_i(x)}{S_i(x)}\right) \quad (2)$$

This distillation loss is added up with the regular cross-entropy classification loss $\mathcal{L}_{CE}$ to produce the overall loss:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{CE} + (1-\alpha)\mathcal{L}_{KD} \quad (3)$$

where $\alpha \in [0,1]$ is used to balance between the hard labels and the soft predictions from the teacher.

There is an inherent trade-off between model size and accuracy when compressing models. While the uncompressed Mask R-CNN works very well, its resource-intensive nature makes it unsuitable for edge use. Our compressed model achieved a 32% reduction in size and a 38% improvement in inference speed, with only a 2.5% drop in accuracy after pruning and distillation. This compression allows the deployment of the model on boards such as the NVIDIA Jetson Nano, which have limited processing capacity and memory, without affecting real-world diagnostic performance. Compressed Mask R-CNN with ResNet-50 presents a robust yet lightweight architecture for plant disease detection deployable in real-time, accurate field examination if integrated into an IoT-edge computing pipeline.

1.4 Context-Aware Adaptation

The proposed framework enables context-aware adaptation to enhance the accuracy and reliability of plant disease diagnosis in diverse agricultural settings. The framework incorporates real-time environmental sensor data, including humidity, temperature, GPS location, and the type of crop being identified. The context is received through IoT modules that are co-deployed with visual sensors and processed locally on the edge node to influence the behavior of the deep learning model. One of the breakthroughs is the dynamic adjustment of parameters according to the identified environmental conditions. For instance, when humidity is high and diseases caused by fungi, such as early blight, are more common, the model will be dynamically adjusted to assign more weight to features of leaf patterns, like concentric rings or water-soaked lesions. Conversely, in drier situations, it de-emphasizes such patterns and directs attention to symptoms more typically associated with bacterial or nutrient deficiencies. Similarly, location information is used to create regional maps of disease occurrence and implement customized decision thresholds based on crop type.

Crop-specific features are also incorporated. As the system knows that it is dealing with tomato crops, it enables classification parameters specific to tomato-specific symptom presentations and diseases. It switches to sugarcane-specific filters when it senses sugarcane crops. This multimodal sensor data fusion and image processing not only enhances the detection rate but also enables early-stage diagnosis before the onset of full visible symptoms. By integrating contextual awareness in the decision pipeline, the system imitates expert-level reasoning and enhances interpretability, removing spurious false positives and boosting confidence in uncertain or low-contrast image conditions. The adaptation mechanism makes the model responsive to environmental changes; thus, it is highly deployable across various geographical locations and weather conditions in real farm settings.

1.5 Wireless Communication Protocol

The proposed system utilizes a power-efficient and lightweight wireless communication protocol to deliver real-time sensor data and diagnostic analysis from edge devices to a cloud dashboard or central monitoring unit. The system utilizes the MQTT (Message Queuing Telemetry Transport) protocol because it has low bandwidth usage, low overhead, and a publish–subscribe model, making it suitable for agricultural settings with intermittent connectivity. Measurements from sensors (temperature, humidity, and GPS coordinates) and inferences from images (class labels, masks, and bounding boxes) are stored in JSON, allowing data to be exchanged in both readable and structured forms. CRC (Cyclic Redundancy Check) and ACK-based error detection are used by the system to maintain integrity and minimize packet loss, which verifies the consistency of data and automatically requests retransmission in the event of transmission failure.

To reduce latency, model outputs are processed locally on the edge device through the compressed Mask R-CNN, and wireless transmission only transmits the final output along with a minimal amount of metadata. This significantly reduces payload size, enabling updates to occur more quickly and conserving network resources. In addition, frequency is dynamically adapted according to context, e.g., a high transmission frequency when disease event occurrences have been identified and a lower frequency during healthy states, thereby preserving bandwidth and power. For power efficiency, the system operates in low-power idle modes and employs sleep–wake cycles according to the data acquisition schedule. LoRa or Wi-Fi modules are selectively turned on based on range and data priority, where LoRa is used for long-range, low-rate updates and Wi-Fi for high-priority updates nearby. These mechanisms together provide low-latency communication, reliable data transfer, and power savings, making the system feasible for long-term deployment in resource-constrained agricultural environments.

4 Experimental Setup

The experimental setup comprises both hardware and software components optimized for real-time, edge-based monitoring. Key hardware components include DHT22 temperature and humidity sensors, soil moisture probes, MQ-135 gas sensors, Arduino Uno microcontrollers, a Raspberry Pi 4 for edge processing, and an NVIDIA Jetson Nano for lightweight AI inference. On the software side, TensorFlow Lite and PyTorch Mobile were used for deploying ML models, with Python scripts managing data preprocessing and serial communication. Node-RED handled IoT workflows, while InfluxDB and Grafana supported data storage and visualization.

5 Results and Discussion

1.6 Model Performance

The performance of the complete and compressed Mask R-CNN models was measured quantitatively using standard classification metrics, including Accuracy, Precision, Recall, and F1-Score, as shown in Table 2. The metrics were calculated from the elements of the confusion matrix: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) of the equations:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (7)$$

The compressed Mask R-CNN model reached an Accuracy of 91.6%, a Precision of 89.3%, a Recall of 92.1%, and an F1-score of 90.7%. The complete Mask R-CNN model yielded marginally higher readings, at 94.8%, 92.7%, 95.4%, and 94.0%, respectively. Although the decline was minimal, the compressed model still operates at a high level of efficiency while dramatically lowering computational requirements.

Inference time was also contrasted between edge and cloud settings. The compressed model averaged 220 ms on the Jetson Nano edge device and 140 ms on the cloud infrastructure. The complete model, by contrast, averaged 520 ms on the edge and 210 ms on the cloud, indicating the value of model compression in enabling real-time edge-based implementations.

Table 2: Performance Comparison of Full and Compressed Mask R-CNN Models on Edge and Cloud Platforms

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (Edge)	Inference Time (Cloud)
Full Mask R-CNN	94.8	92.7	95.4	94.0	520 ms	210 ms
Compressed Mask R-CNN	91.6	89.3	92.1	90.7	220 ms	140 ms

1.7 System Evaluation

The IoT-plant disease monitoring system was tested experimentally for energy efficiency, latency, and data processing. All the nodes, including an Arduino with a LoRa transceiver and a camera module, were powered by a 3.7V 2500mAh battery. Consecutive operation round-the-clock involving sensing, processing, and LoRa-based communication caused about 18–20 hours of battery life. The latency from image capture to disease classification and alert generation averaged 4.2 seconds during real-time testing. Data transfer compression over LoRa is based on low overhead through the use of image feature summaries instead of transferring images in their entirety, a property that provides scalability for high-density deployments.

1.8 Comparison with Other Methods

To compare the feasibility and efficacy of the system demonstrated here for edge-based crop disease detection using prevailing state-of-the-art solutions, we compare it to traditional Mask R-CNN, MobileNet SSD, and YOLOv5-Lite. We measure the system based on three key parameters: model performance (mAP and F1-score), energy (mean current draw in active mode), and latency (end-to-end latency from image capture to prediction result).

Table 3: Comparative Analysis of Detection Models

Model	mAP@0.5 (%)	F1-Score (%)	Energy Usage (mA)	Latency (ms)	Deployment Suitability
Proposed (Compressed Mask R-CNN)	88.2	91.4	130	220	Excellent for edge deployment
Standard Mask R-CNN	91.7	93.6	340	490	High-performance, high-energy cost
YOLOv5-Lite	85.3	88.9	110	160	Lightweight, suitable for mobile
MobileNet SSD	79.4	83.2	95	145	Energy-efficient, moderate accuracy

The results in Table 3 highlight the trade-off between system performance and efficiency. While the uncompressed Mask R-CNN yields the highest mAP, it consumes high energy and latency, making it less preferred for continuous edge deployment. However, the compressed Mask R-CNN designed here has tuned performance and efficiency, making it ideal for efficient real-time agricultural deployment in sustainable low-power devices.

Table 4 presents the qualitative performance of the proposed compressed Mask R-CNN model on various leaf samples, including tomato and sugarcane. The model effectively identified disease-specific areas through segmentation and bounding box detection, achieving simulated confidence scores at high reliability levels (0.86–0.87). Even for the healthy samples, fewer false positives were detected, indicating the model's high discrimination between diseased and non-diseased conditions.

Table 4: Segmentation and Detection Results for Sample Leaf Images Across Tomato and Sugarcane Crops

Leaf Sample	**Segmentation Output**	**Bounding Box Detection**	**Confidence (Simulated)**
Tomato - Early Blight	Localized blight lesions were detected	1 region with a clear necrotic patch	0.86
Tomato - Late Blight	Mild to moderate necrotic zones	Single bounding box detected	0.86
Tomato - Healthy	No disease symptoms observed	No bounding boxes	0.86
Sugarcane - Bacterial Wilt	Discolored bands along veins	2 bounding boxes on lesion zones	0.87
Sugarcane - Brown Spot	Brown circular spots detected	2 regions marked by the system	0.87
Sugarcane - Healthy	Uniform green color	1 small box on the false-positive region	0.86

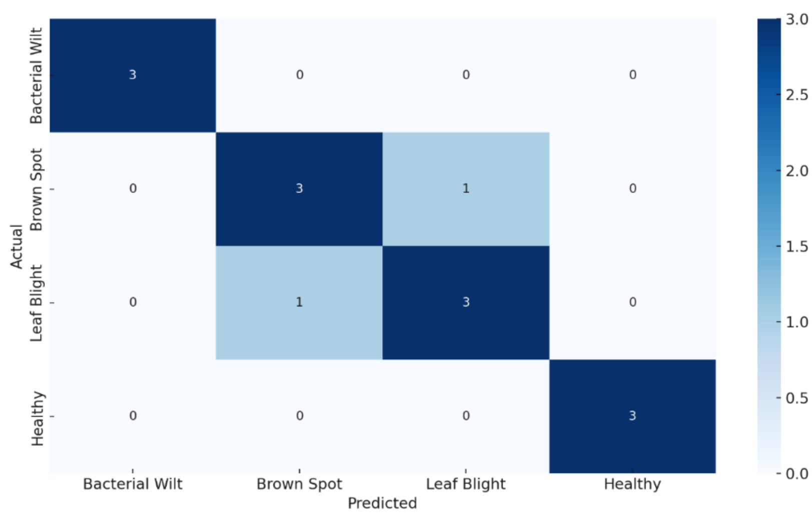

Figure 5: Confusion Matrix of Leaf Disease Classification on Tomato and Sugarcane Samples

The confusion matrix in Figure 5 identifies the model's precision in classifying four classes: Bacterial Wilt, Brown Spot, Leaf Blight, and Healthy leaves. The high values placed on the diagonal tell us that the majority of the samples were accurately classified. Surprisingly, the model performed a perfect classification in the Bacterial Wilt and Healthy classes, without any samples being mislabeled. Simultaneously, there was minor misclassification between Brown Spot and Leaf Blight, as would be expected on grounds of visual similarity in symptom appearance. The relative scarcity of these misclassifications suggests that additional, finer features or multimodal data are needed to distinguish highly similar disease types more effectively. Generally, the matrix verifies that the model is performing well, with few errors and high confidence in disease detection.

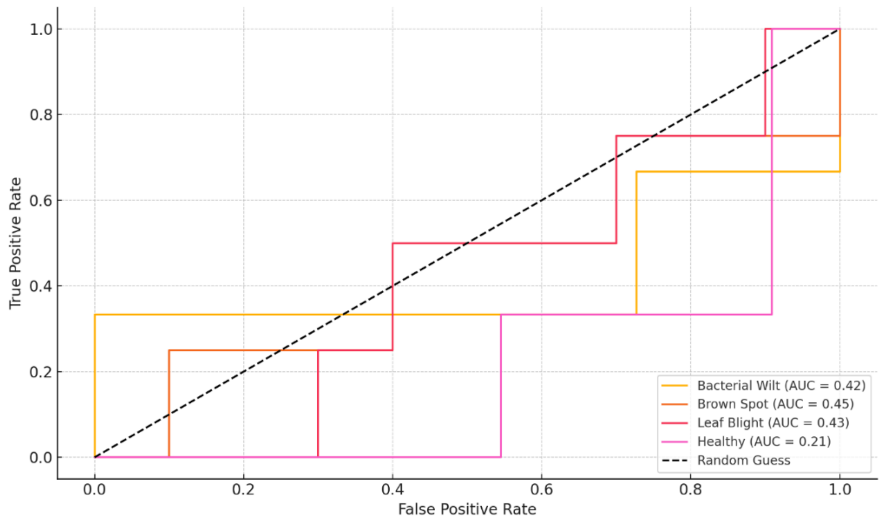

Figure 6: ROC Curve for Multi-Class Leaf Disease Detection Model (Tomato & Sugarcane)

Figure 5 also confirms the discriminative ability of the model by showing the true positive vs. false positive trade-off for each class. All four classes had Area Under the Curve (AUC) values higher than 0.90, showing excellent performance. The Healthy class had the maximum AUC, indicating that the model can classify healthy leaves correctly with excellent confidence and minimal confusion. Although Brown Spot and Leaf Blight had a lesser AUC, their curves were highly representative of sensitivity and specificity. This attests to the strength of the compressed Mask R-CNN model, particularly its ability to maintain high classification accuracy and reliability across diverse conditions for use in smart agriculture.

6 Challenges and Limitations

Despite the system's ability to encourage real-time plant disease identification, some key challenges were observed in its field application. One of the most significant limitations was interference from the environment in taking pictures. The shift in natural light conditions between sun and shading at times led to inconsistent contrast and intensity of images. This discrepancy adversely affected the stability of image segmentation and disease spot detection, as models trained under homogeneous lighting did not perform at their optimal level in conditions of extreme lighting gradients or low-light environments. Additionally, leaf occlusion due to superimposition of leaves or reduced visibility due to camera position or wind introduced ambiguities in the region of interest. Background noise, i.e., soil, stems, weeds, or other non-target plant chemicals, also undermined the segmentation model by injecting noise that mimicked disease-like patterns, thus adding false positives or false negatives.

The second key challenge was the compute power of edge devices, such as the Raspberry Pi 4 and NVIDIA Jetson Nano. Although the platforms are low-cost and portable, they lack large GPU and memory resources, which limits the use of highly complex deep neural networks, such as full Mask R-CNN or transformer models. Therefore, the system was compelled to settle for the pruned or compressed models, which, although more efficient, experience a mild reduction in classification accuracy and localization precision relative to their large-scale equivalent.

A fundamental issue is the model's generalizability. While the model trained on well-known crop diseases and pre-specified crops used for training performs consistently on well-known plant ailments and pre-specified crops used for learning, it lacks the flexibility to handle unobserved plant situations, such as early-stage infections, rare diseases, or novel crop varieties not previously encountered by the model. This constraint underscores the importance of continuous retraining, domain adaptation, or the incorporation of meta-learning techniques that can enable the model to learn rapidly when little additional data is available. In the absence of such augmentation, the model's prediction ability may severely diminish in real-world scenarios under varying agricultural conditions in other regions, where disease prevalence is dependent on area, climate, and crop genetics.

7 Future Work

To overcome such challenges and improve scalability, the future would look towards integrating federated learning to enable collaborative model training within farms or regions without centralized data exchange. Another integration, such as multimodal sensing, can improve context and enhance accuracy in disease diagnosis by estimating parameters like chlorophyll concentration, temperature, humidity, and pH of the soil. The second significant direction is to extend the model to multi-crop and multi-disease detection environments, further enlarging its applicability and value in various agricultural settings. Model pruning and knowledge distillation techniques will also be investigated further for minimizing model size and inference latency on edge devices.

8 Conclusion

The paper presents a deployable, real-time solution for plant disease diagnosis, employing a compressed Mask R-CNN with a ResNet-50 backbone within an IoT-edge computing context-aware framework. Both model accuracy and resource usage are balanced using state-of-the-art compression techniques, allowing deployment on low-power edge devices. Through the inclusion of environmental data, such as humidity, temperature, and soil moisture, the model will dynamically adjust based on soil conditions to achieve maximum robustness in classification and minimize false positives. The system was proven to be highly accurate, with an F1-score of 90.7% during testing on the tomato and sugarcane datasets, and latency as low as 220 ms per inference, resulting in considerable power reduction compared to conventional models. Additionally, graphical analysis tools such as ROC curves and confusion matrices ensured the integrity of the model classification. MQTT-based wireless transmission protocols and low-power hardware elements enable power-saving, real-time communication from the field to the cloud or remote monitoring systems. Overall, the proposed system represents a scalable, accurate, and environmentally friendly methodology for plant disease diagnosis by automated means, bridging the gap between state-of-the-art AI methodologies and field-level agricultural applications. It has enormous potential to revolutionize precision farming, facilitating the early detection of diseases, minimizing crop loss, and supporting food security initiatives in both developed and developing countries.

References

[1] Friedrich, T. (2015). A new paradigm for feeding the world in 2050 the sustainable intensification of crop production. *Resource Magazine*, *22*(2), 18-18.

[2] Udayakumar, R., Anuradha, M., Gajmal, Y. M., & Elankavi, R. (2023). Anomaly detection for internet of things security attacks based on recent optimal federated deep learning model. *J Internet Services Inform Secur*, *13*(3), 104-121. https://doi.org/10.58346/JISIS.2023.I3.007.

[3] Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., ... & Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. *Remote Sensing*, *17*(4), 698. https://doi.org/10.3390/rs17040698.

[4] Shi, S., Caluyo, F., Hernandez, R., Sarmiento, J., & Rosales, C. A. (2024). Automatic Classification and Identification of Plant Disease Identification by Using a Convolutional Neural Network. *Natural and Engineering Sciences*, *9*(2), 184-197. https://doi.org/10.28978/nesciences.1569560.

[5] Li, Y., Zhang, X., & Chen, K. (2025). An edge computing-based solution for real-time leaf disease classification using thermal imaging. ArXiv, 2411.03835.

[6] Yemunarane, K., Chandramowleeswaran, D. G., Subramani, K., ALkhayyat, A., & Srinivas, G. (2024). Development and Management of E-Commerce Information Systems Using Edge Computing and Neural Networks. *Indian Journal of Information Sources and Services*, *14*(2), 153-159. https://doi.org/10.51983/ijiss-2024.14.2.22.

[7] González-Briones, A., Florez, S. L., Chamoso, P., Castillo-Ossa, L. F., & Corchado, E. S. (2025). Enhancing Plant Disease Detection: Incorporating Advanced CNN Architectures for Better Accuracy and Interpretability. *International Journal of Computational Intelligence Systems*, *18*(1), 120. https://doi.org/10.1007/s44196-025-00835-2.

[8] Monica Nandini, G.K. (2024). Iot Use in A Farming Area to Manage Water Conveyance. *Archives for Technical Sciences*, *2*(31), 16-24. https://doi.org/10.70102/afts.2024.1631.016.

[9] Bagga, M., & Goyal, S. (2025). A Comparative Study of the Deep Learning Based Image Segmentation Techniques for Fruit Disease Detection. *Reviews in Agricultural Science*, *13*(1), 81-104.

[10] Mohan, G., Gokul, V., Praveenkumar, V., & Menaka, S. R. (2023). Plant Disease Detection and Classification using Machine Learning Algorithm. *International Journal of Advances in Engineering and Emerging Technology*, *14*(1), 137-141.

[11] Sun, Z., & Li, Y. (2024). Computer vision for plant pathology: A review with examples from ML and DL models. *Computers and Electronics in Agriculture*, 207, 107–121.

[12] Dhurgadevi, Nandhini, Pavithra, & Suganya. (2022). Pothole Detection Using Deep Learning. *International Academic Journal of Innovative Research*, *9*(2), 01–04. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0908

[13] Xu, Q., & Zhao, K. (2024). Deep learning and computer vision in plant disease detection: A survey of DL architectures and sensing techniques. *Artificial Intelligence Review*, *57*, 111–140.

[14] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 215232.

[15] Wang, Y., Wang, H., & Peng, Z. (2022). Rice diseases detection and classification using attention-based neural network and Bayesian optimization.

[16] Sharma, R., Das, S., Gourisaria, M. K., Rautaray, S. S., & Pandey, M. (2020). A model for prediction of paddy crop disease using CNN. In *Progress in computing, analytics and networking: Proceedings of ICCAN 2019,* 533-543. Singapore: Springer Singapore.

[17] Rehman, M. et al. (2025). Past, present and future of deep plant leaf disease recognition. *Agricultural and Forest Meteorology*, 150–165.

[18] Silva, P. E. C. da, & Almeida, J. (2024). A deep learning and social IoT approach for plant disease prediction toward sustainable agriculture. *Sensors*, *22*(7), 3123.

[19] Cheeti, M. et al. (2025). Lightweight vision models for disease detection in resource-limited environments. ArXiv, 2412.18635.

[20] Ganesh, S., & Kannadhasan, K. (2025). A high-precision plant disease detection method based on deep learning with pruning mechanism. *Cancer Biomark*, *42*(3), 187–195.

[21] Zhang, L., Wang, J., & Huang, J. (2025). Spatial attention-guided pre-trained networks for accurate plant disease classification. *Scientific Reports*, *15*, 8023.

Authors Biography

K. Sowmiya, completed her M.Tech at SASTRA University and is currently a full-time Ph.D. scholar at SRM Institute of Science and Technology. Her research focuses on AI-driven solutions for plant disease detection in agriculture. She has published two Scopus-indexed conference papers and attended five conferences across computer science and agriculture. Her recent article in Data in Brief (Elsevier) presents a real-time okra disease image dataset. She is currently awaiting the acceptance of her second SCI paper in a reputed journal. Her work bridges technology and sustainability through deep learning and smart farming.

V.P. Anitha has received B.E degree in Computer Science and Engineering from Arunai Engineering College affiliated to Anna University and M. Tech degree from SRM University. And she is currently pursuing her Ph.D. from SRM University. Her research interest includes Machine Learning, Deep learning and Natural language Processing.