Attention-Fused Residual Lightweight MobileViT for Plant Leaf Disease Detection in Agriculture

Layth Hussein[1]*

[1]Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq

Abstract. In recent years, early detection and management of plant diseases at an early stage has become very important for sustaining crop health, increasing yields, and protecting food security in agricultural systems. Untreated diseases caused by fungi, bacteria, viruses, and pests significantly reduce agriculture and pose a risk to global food production. However, existing models face some challenges in agriculture because they focus only on local features, fail to focus on noisy backgrounds, lighting, overlapping leaves, and difficulty in detecting different parts of the leaf. To overcome this problem, this paper proposes an attention-fused residual lightweight (RTR-Lite-MobileViT) model, which is an enhanced version of the original MobileVit model designed for efficient deployment on resource-constrained devices. Different attention techniques, such as Squeeze-and-Excitation Networks (SENet), Efficient Channel Attention (ECA) and Triplet Attention, are added to reduce the computational footprint of the model while boosting its ability to capture complicated disease patterns. Experiments were validated using publicly available datasets, such as the Plant Disease and PlantDoc datasets. Next, the performance of proposed RTR-Lite-MobileViT, obtained high results of accuracy (99.95%), precision (98.97%) in plant disease dataset and accuracy (99.96%), precision (97.98%) in Plantdoc dataset compared to existing model like MobileNetV2.

1 Introduction

In recent years, plant leaf disease detection has attracted substantial attention as a vital aspect of precision agriculture. Plant diseases caused by fungi and bacteria contribute to yield losses of up to 40%, threatening global food security and farmers' livelihoods [1]. However, traditional manual inspection methods are time-consuming, error-prone, and require practical knowledge, and are often inaccessible in remote farming regions. Therefore, the incorporation of Deep Learning (DL) and computer vision has been developed as a solution for early, automated, and consistent disease identification. These techniques permit efficient monitoring of crop health, timely intervention, and minimization of excessive pesticide use. Consequently, plant disease detection using Artificial Intelligence (AI) and lightweight DL models is becoming a keystone for attaining sustainability and resilience in current agriculture [2]. Despite these advancements, plant leaf disease detection faces various challenges, including generalization across various conditions because real-world environments vary in lighting, background complexity, leaf maturity, and disease progression. Many models trained on curated datasets fail to successfully adapt to these variations [3]. Simultaneously, lightweight yet accurate models are designed. While DL models attain greater accuracy, their large parameter sizes and computational requirements make deployment harder on resource-constrained devices such as Internet of Things (IoT) sensors or mobile platforms. Overcoming these limitations requires strong adaptive architectures that can function in dynamic agricultural environments while maintaining performance and reliability [4]. Moreover, DL models attain high accuracy in plant disease detection, but are computationally expensive, which limits their use in mobile or IoT devices [5]. Lightweight networks often fail to capture complex disease features across various crops. Incorporating advanced attention mechanisms such as the Squeeze-and-Excitation Network (SENet) and Triplet Attention supports feature representation, enhancing the classification accuracy and robustness while keeping models computationally feasible for real-time agricultural deployment [6,7]. The state-of-the-art methods include the Lightweight Ensemble Model of MobileNetV2 and Xception (LEMoxinet). In this method, MobileNetV2 is employed for efficient computation on resource-limited devices, although the Xception network is utilized for its powerful feature extraction capability. The ensemble of these two models improves the classification accuracy and robustness across several plant disease classes [7]. Moreover, the model can achieve a strong balance between the efficiency and accuracy. However, the ensemble significantly increases the overall model complexity and memory requirements, making it less appropriate for real-time deployment on low-power agricultural devices [8]. Simultaneously, the Hybrid Residual Network-50 with a convolutional block attention module and support vector machine (SVM). In this method, Residual

* Corresponding author: laith.h.alzubaidi@gmail.com

Network-50 performs feature extraction, the convolutional block attention module highlights the most applicable disease-affected regions in the leaf images, and the SVM delivers the final classification [9]. This combination improves interpretability and classification accuracy. The advantage is the more precise detection of critical disease features [10]. However, the model struggled with its increased computational cost and inference time, which decreased its practicality for mobile or IoT-based agricultural platforms [11].

The key contributions of the research are:

- The Novel Lightweight Hybrid Architecture named RTR-Lite-MobileViT is proposed which is an enhanced lightweight vision transformer that combines CNN-based local feature extraction with transformer-based global reasoning for accurate and robust plant disease detection.
- A multilevel attention mechanism combining the Efficient Channel Attention (ECA) Triplet Attention is introduced to emphasize disease-relevant regions while suppressing background noise, illumination artifacts, and overlapping leaf interference.
- Residual Lightweight Optimization is utilized to ensure deployment feasibility on resource-constrained devices, by which the framework incorporates Depthwise Separable Convolutions and residual shortcuts that assist in significantly reducing computational complexity and parameter count while preserving recognition accuracy.

The overall research is structured as follows: Section 2 presents a literature review, Section 3 defines the proposed Methodology, Section 4 presents the Experimental Results, and Section 5 concludes the paper.

2 Literature review

Sangeetha Duhan et al [12]. demonstrated a lightweight plant-disease detection and classification model. Initially, data were sourced from several datasets, such as the plant disease dataset, coffee, and paddyDocter, which contains pictures of various crop diseases from diverse geographical locations. The first data source was preprocessed using data augmentation and image resampling, where data augmentation for image flipping to a 30 °angle and image resampling was applied to resize all the images to 224×224 pixel size. Later, preprocessed data were fed into the RTR_Lite_Mobilenet model, where different techniques were applied, such as squeeze-and-excitation networks (SENet) and triplet attention. Therefore, these techniques enhance model performance by reducing model computation and increasing the ability to identify complex plant diseases effectively. Moreover, this model has the potential to alert farmers by processing crop images when disease is detected, which leads to saving energy and bandwidth networks. However, this model struggles to adapt to real-life field testing, particularly in dynamic agricultural environments.

Masoud Rezaei et al [13]. developed a plant disease recognition system using a few-shot learning (FSL) technique in low-data scenarios. Initially, data were collected from two publicly available datasets, namely, the PlantVillage dataset and PlantDoc dataset, where the PlantVillage dataset consists of 54309 images of healthy and non-healthy leaves, and the PlantDoc dataset contains 2598 images of 13 plant species. The first collected data were preprocessed using normalization and data augmentation, where normalization was applied for resizing pixels to the same size and data augmentation to flip horizontal images, change contrast, brightness, and saturation. Later preprocessed data was fed into FSL pipeline where it combines Pre- training, meta-learning and Fine tuning (PMF) along with these methods it uses Feature Attention (FA). Therefore, PMF + FA effectively reduces the complexity of the background of the image and makes the model understand the objects clearly. Moreover, this model can capture various crop diseases in distinct geometric regions. However, this model struggles to identify diseases, particularly if the images are blurred.

Emmanuel Moupojou et al [14]. introduced a plant-disease detection and segmentation system for field images. Initially, data were sourced from publicly available datasets, namely the PlantVillage dataset, PlantDoc, FieldPlant, and PlantDisease 271 datasets, which contain different crop disease images, vegetables, grains, trees, and plants. The first collected data were preprocessed using various techniques to identify hidden objects from all images. The preprocessed data were then fed into a Convolutional Neural Network (CNN) based on MobileNetV2 processes using ImageNet pretrained weights and further trained with Fully Convolutional Data Description (FCDD) for outlier detection, which identifies all anomalies. Moreover, this model can accurately adapt when new diseases are found in various crops. However, this model failed to perform accurately under different weather conditions.

Xue Jiang et al. [15] presented PlantCaFo, an efficient few-shot plant disease recognition framework that highlights the foundation models for enhanced performance in data-scarce agricultural scenarios. Initially, experiments were directed on the PlantVillage and Cassava datasets, followed by the structure of an out-of-distribution dataset (PDL), which consisted of 21 categories to estimate generalization. Then, the processed data were fed into (DCon-Adapter) to improve global and local feature learning and a Weight Decomposition Matrix (WDM) to adjust image-text alignment with minimal parameters. Additionally, data augmentation techniques are applied, such as Mixup and CutMix, to increase the strength of the limited training samples. Finally, the model attained 93.53% accuracy in a “38-way 16-shot” setting over baselines on real-world cassava images. Moreover, the model is capable of greater generalization with fewer samples. However, the performance of the model is reduced in highly varied and complex real-world field situations.

Salman et al. [16] established a novel approach to plant disease detection under uncontrolled field conditions. Initially, data were collected from the PlantVillage (lab-controlled images) and PlantDoc (real-world in-the-wild images) datasets. The collected data were then preprocessed by down sampling using t-distributed Stochastic Neighbor Embedding (t-SNE) clustering with Visual Geometry Group 16-layer network (VGG16) features to overcome class imbalance and redundancy. Images were then fed into a hybrid model combining a Vision Transformer (ViT) backbone for global feature extraction and a Mixture of Experts (MoE) with a gating mechanism for specialized classification. Finally, the model achieved better results for PlantVillage, which significantly outperformed the baselines. Moreover, the model exhibited robust generalization in diverse situations. However, the model presented a high computational overhead owing to numerous experts and gating, which limits the deployment of resource-constrained devices.

The residual attention-fused lightweight mobile vision transformer (RTR-Lite-MobileViT) was designed as an optimized hybrid architecture for computational efficiency and accuracy in agricultural disease detection tasks. Most traditional CNN-based approaches focus on local features and not on context which makes them more sensitive to noise from the background, illumination variability, and occlusion from the overlapping leaves of plants the most common vehicles for damage caused by diseased plants. While the ViT architecture considers a global context, ViTs are computationally heavy. Specifically, RTR-Lite-MobileViT has MobileViT as its backbone and multilevel attention fusion (SE, ECA, Triplet Attention) with residual lightweight modules utilizing Depthwise Separable Convolutions for increased discriminative power while remaining suitable for resource-constrained agricultural devices such as drones, smartphones, and edge IoT or edge device systems. Figure 1 shows flow diagram of proposed RTR-Lite-MobileViT model.

3 Methodology

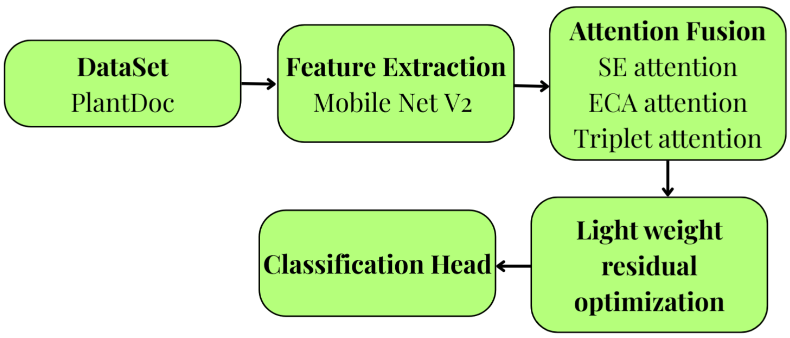

Fig. 1. Flow diagram of proposed RTR-Lite-MobileViT model.

3.1 Dataset description

The PlantDoc dataset [14] is a realistic benchmark for plant disease identification, with approximately 2,598 images across 13 species of plants (mostly tree species) and 27 classes of diseases, whose categories include healthy samples of plants. Unlike other laboratory-controlled datasets, images in this dataset were all taken under field conditions while observing various impediments to plant health, such as cluttered backgrounds, unusual lighting, overlapping leaves, and various orientations. Hence, PlantDoc datasets are difficult to distinguish and are used to assess the robustness of plant disease identification models.

3.2 Proposed RTR-Lite-MobileViT

In disease detection, local details, such as leaf spots and edge discoloration, provide the first indicators of infection, while global structures include the spread of lesions across the leaf surface that determine disease severity. In particular, CNNs capture local patterns effectively but fail to integrate long-range dependencies, whereas ViTs capture global structures, but require high computational power. MobileViT bridges this gap by combining CNN-based local feature extraction with ViT-based global reasoning in a lightweight design.

The input leaf image is defined as $I \in \mathbb{R}^{H \times W \times 3}$, and the local feature maps are illustrated in Equation (1).

$$F_{local} = Conv(I), F_{local} \in \mathbb{R}^{h \times w \times c} \quad (1)$$

$I \in \mathbb{R}^{H*W*3}$ denotes the input RGB leaf image, where H and W are the image height and width, respectively. $F_{local} \in \mathbb{R}^{h*w*c}$ illustrates the local feature maps extracted by convolutional layers, h ω are reduced spatial dimensions, and c is the number of channels.

Further, these local embeddings are reshaped into patch sequences using Equation (2):

$$X = \text{Unfold}(F_{\text{local}}), X \in \mathbb{R}^{N \times d} \quad (2)$$

$X \in \mathbb{R}^{N*d}$ denotes Sequence of N patch embeddings with embedding dimension d.

Furthermore, a transformer encoder models the global dependencies, as shown in Equation (3):

$$F_{\text{global}} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (3)$$

F_{global} signifies Global feature representation obtained via transformer encoder.

Finally, both levels are fused through the Equation (4):

$$F = \text{Fusion}(F_{\text{local}}, F_{\text{global}}) \quad (4)$$

F denotes a fused feature map combining the local (F_{local}) and global (F_{global}) information. F_{att} illustrated the attention-enhanced feature map after applying SE, ECAA and Triplet attention. F_{res} denotes the final residual-enhanced feature map after lightweight residual connections. This ensures that tiny lesions and larger-scale spread patterns are jointly represented, which assists in enhancing the disease recognition.

3.3 Attention fusion

Images of agricultural products often contain background distractions (e.g., dirt, stalks, sky), and inconsistent lighting can obscure subtle cues of disease symptoms. Attention mechanisms can be considered as a spotlight that allows the model to identify regions of the leaf that are affected by disease while ignoring unrelated noise. The proposed RTR-lite-MobileViT method uses three complementary attention modules, as explained below.

3.3.1 Squeeze-and-excitation (SE)

The SE orients the discriminative channels in the spectrum by leveraging global channel dependencies. This is particularly important because the cues of a diseased plant often manifest vastly differently in the spectral response (e.g., yellow chlorosis vs. brown necrosis), as demonstrated in Equation (5):

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j), s_c = \sigma(W_2 \delta(W_1 z_c)) \quad (5)$$

Where z_c signifies the channel descriptor for channel c_t obtained via global average pooling. S_c represents the channel attention weight for channel c in the SE block. W_1, W_2 denote the trainable weight matrices of the fully connected layers in the SE block. $\delta(.)$ where represents the ReLU activation function. $\sigma(.)$ where is the sigmoid activation function. C illustrates the number of feature channels.

Hence, this boosts channels that encode disease symptoms while suppressing irrelevant features.

3.3.2 Efficient channel attention (ECA)

SE assists in improving accuracy but is computationally heavy. For edge deployment, an ECA is utilized because it provides a lighter solution by using a 1D convolution to capture channel interactions, as demonstrated in Equation (6):

$$k = \left| \frac{\log_2(C)}{\gamma} + b \right|_{\text{odd}}, \alpha = \text{Conv1}D_k(z) \quad (6)$$

Here, k denote the convolutional kernel size in the ECA block, adaptively determined by C. γ, b denotes the hyperparameters for the kernel size computation in ECA. α denotes the channel weights computed by the ECA via a 1D convolution. F_p Feature map rotated along the dimension $p \in \{HW, HC, WC\}$ for Triplet Attention. Thus, the employed ECA helps balance accuracy and efficiency, which ensures the effective analysis of smartphones or IoT devices.

3.3.3 Triplet attention

Specifically, diseases often manifest as irregular shapes across different leaf orientations. Further, triplet attention captures cross-dimensional dependencies by rotating tensors across three planes (spatial and channel), as illustrated in Equation (7):

$$F_{att} = \frac{1}{3} \sum_{p \in \{HW, HC, WC\}} \text{Attention}(F_p) \quad (7)$$

Hence, this helps the model to detect lesions consistently, even when the leaves are tilted, overlapped, or partially occluded.

The fused attention-enhanced features are formulated as expressed in Equation (8).

$$Fatt = Triplet(ECA(SE(F))) \quad (8)$$

Detection systems in the agricultural sector are frequently instantiated in drones, handhelds, or IoT nodes, where memory, processing, and energy capabilities are limited. Heavy models can hinder real-time decision-making. To resolve this issue, RTR-Lite-MobileViT utilizes Depthwise Separable Convolutions (DsConv), which help to reduce this complexity and therefore uses residual shortcuts.

3.3.4 Depthwise separable convolution (DsConv)

The conventional convolution is computationally intensive. Furthermore, the convolution is decomposed by Ds, in which depth-wise and point-wise are illustrated in Equation (9):

$$Params_{Conv} = K^2 MN, Params_{DsConv} = K^2 M + MN \quad (9)$$

Where K denotes the size of the convolutional kernel (e.g., 3 × 3). M represents the number of input channels for a convolutional layer. N denotes the number of output channels in the convolutional layer. $Params_{Conv}$ represents the number of parameters in a standard convolution. $Params_{DsConv}$ denotes the number of parameters in depth-wise separable convolution. Hence, this simplifies the work by almost an order of magnitude (MobileViT), which allows an effective operation with no reduction in accuracy.

3.3.5 Residual shortcuts

Residual learning preserves low-level lesion textures such that they may still be picked up in the deeper layers of the network. This is important because some diseases demonstrate both micro symptoms (small spots) and macro symptoms (across the veins). As illustrated in Equation (10),

$$F_{res} = F_{att} + \mathcal{H}(F_{att}) \tag{10}$$

$\mathcal{H}(.)$ illustrates the transformation function (e.g., DsConv or MobileViT block) applied to the residual connection. Hence, the proposed RTR-lite-MobileViT framework encompasses feature learning in both local and global contexts, multilevel blended attention, and lightweight residual learning in an end-to-end training framework, specifically for agricultural diseases and crop inspection. Our model can leverage MobileViT's hybrid representation capabilities with SE, ECA and Triplet attention fusion to classify the relevant area of disease while minimizing unwanted features and noise caused by complex crop field environments. Depthwise separable convolutions and residual shortcuts offer computational efficiency and make it easier to deploy on mobile and edge devices in real time. In addition, our training strategy and evaluation measures allow our model to be robust and generalizable using a variety of agricultural datasets. Thus, this methodology is promising for accurate, efficient, and scalable plant disease detection in real-world agricultural use cases.

4 Experimental results

In this study, an efficient RTR-Lite-MobileViT model was proposed for detecting and classifying plant diseases. The RTR-Lite-MobileViT model was used to validate and detect real-time plant diseases. The proposed model was implemented on Windows 10 Pro with an Intel i3 processor running at a speed of 3.6GHz. The performance evaluation metrics include accuracy, precision, recall, F1-score, TPR, and FPR, which are expressed in Equations (11)-(14):

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \tag{11}$$

$$Precision = \frac{TP}{FP+TP} \tag{12}$$

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

$$f1score = \frac{(2*precision*recall)}{(precision+recall)} \tag{14}$$

Here, TP and TN denote the true positive and true negatives, respectively. FN and FP denote the false positives and false negatives, respectively.

4.1 Performance analysis

The performance of the proposed RTR-Lite-MobileViT model was compared with other models such as InceptionResnetV2 and Vision transformer, and the performance of these models was low when compared with the performance of the RTR-Lite-MobileViT model. The proposed model is the one model has significantly improved accuracy when compared with all other mentioned models.

Table 1. Performance analysis of proposed RTR-Lite-MobileViT model.

Performed Models	Accuracy (%)	Precision	Recall
InceptionResnetV2	97.52	97.34	97.62
Vision transformer (ViT)	98.54	98.67	98.23
Proposed RTR-Lite-MobileViT	99.92	99.87	99.85

Table 1 shows Performance analysis of proposed RTR-Lite-MobileViT model. Traditional models, such as InceptionResnetV2, Vision transformer (ViT), and the proposed RTR Lite MobileNetViT model, attained superior accuracies of 97.52, 98.54, and 99.92, respectively.

4.2 Comparative analysis

The proposed RTR-Lite-MobileViT model model was compared with other existing models, such as Vision Transformers, and the model was also used for the detection of plant diseases and got evaluated. The proposed model was compared with Vision Transformers.

Table 2. Comparative analysis of proposed RTR-Lite-MobileViT.

Performed Models	Accuracy (%)	Precision	Recall
ViT	86.79	94.07	94.56
Proposed RTR-Lite-MobileViT model	99.92	99.87	99.85

The Table 2 above shows a detailed analysis of the performed model along with its traditional model and its accuracy. The proposed model achieved an accuracy of 99.92 and the traditional vision transformer model achieved an accuracy of 86.79. Additionally, the proposed RTR Lite MobileNet ViT model uses a triplet attention mechanism that arrests cross-dimensional interactions with minimal computation.

4.3 Discussion

The experimental results demonstrated that the RTR-Lite-MobileViT model could resolve key issues present in the state-of-the-art model for agricultural disease detection. Existing lightweight CNN models, such as MobileNet and EfficientNet, only encapsulate local texture features that are susceptible to interference from complex background information, constitutive illumination variation, or duplication of occluded leaves. Novel architectures or transforms often highlight the ability to recognize global dependencies. However, such large-scale and equally inefficient architectures have serious deployment costs, requiring us to rely on complex architectures while maintaining reasonable accuracy. The RTR-Lite-MobileViT architecture optimizes these challenges with the decomposition of clarity, mobile viability, multi-attention fusion, and the low residual efficiency of MobileViT. Resource utilization is shown in many types of improved performance over state-of-the-art models using Plant Disease and PlantDoc datasets. The introduction of SE and ECA modules highlights electronic indices from representational similarities that become paramount in isolated disease features emerging among tens of leaves. Triplet Attention and included attention to identifying the background between structures leaves to create separation in representation, especially to compensate irrelevant leaf features, is critical for definitions of subtle disease-related lesions largely unaccounted for in many conventional architectures.

The use of Depthwise Separable Convolutions as an attempted scheme maintains a lower computational burden to acquire lighter architectures, as shown in RTR-Lite-MobileViT, while maintaining similar representation power and more available inference time usage in comparison to their larger transformer variants. Despite advancements, there are still some challenges. The model has the ability to handle issues with background noise and illumination, but its performance may be impacted when disease symptoms are very subtle in nature or when they are visually similar across a mixture of classes. The model also has a bias to image features alone and therefore cannot incorporate multimodal information, such as spectral indices, soil characteristics, or weather conditions. These factors may need to be incorporated into overall disease diagnosis. Finally, another limitation may be the confusion of visually similar disease classes, which may require more complex disentangling of the features themselves.

5 Conclusion

The proposed RTR-Lite-MobileViT is an attention-fused residual lightweight DL model for effective plant disease detection in actual agricultural circumstances. Specifically, the proposed architectural structure includes MobileViT as a backbone with three multi-level attentions that consist of SE, ECA, Triplet attention, and lightweight residual optimization utilizing depth-wise separable convolutions. Furthermore, the framework captures local lesion texture information with global information concerning the whole plant or crop while eliminating background clutter and confounding/illuminating artifacts. Thus, the experimental results were validated on both the Plant Disease and PlantDoc datasets, which demonstrated that the proposed RTR-Lite-MobileViT outperformed the conventional CNN and transformer-based architecture with greater precision, recall, and F1-scores while handling low computational cost for mobile and IoT deployments, where the proposed model achieved an accuracy of 99.92. Thus, the proposed RTR-site-MobileViT is an accurate, reliable, and scalable system for plant disease detection. In the future, this research will be enhanced by narrowing the performance-to-deployment gap for next-generation agricultural-based high-performance deep learning models for plant disease detection.

References

1. Y. Miao, W. Meng, X. Zhou, SerpensGate-YOLOv8: An enhanced YOLOv8 model for accurate plant disease detection. Front. Plant Sci. **15**, 1514832 (2025). https://doi.org/10.3389/fpls.2025.1514832
2. Y. Wang, P. Zhang, S. Tian, Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. Front. Plant Sci. **15**, 1382802 (2024). https://doi.org/10.3389/fpls.2024.1382802
3. K. Joshi, S. Hooda, A. Sharma, H. Sonah, R. Deshmukh, N. Tuteja, S.S. Gill, R. Gill, Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation. Curr. Plant Biol. **41**, 100437 (2025). https://doi.org/10.1016/j.cpb.2025.100437
4. R. Mao, Y. Zhang, Z. Wang, X. Hao, T. Zhu, S. Gao, X. Hu, DAE-Mask: A novel deep-learning-based automatic detection model for in-field wheat diseases. Precis. Agric. **25**, 785-810 (2024). https://doi.org/10.1007/s11119-023-10093-x
5. P. Indumathi, R. Kumuthaveni, Coati optimized transfer learning with vision transformer model for improving deep learner based plant diseases detection. Int. J. Intell. Eng. Syst. **17**, 277-288 (2024). https://doi.org/10.22266/ijies2024.0831.21
6. A. Singla, A. Nehra, K. Joshi, A. Kumar, N. Tuteja, R.K. Varshney, S.S. Gill, R. Gill, Exploration of machine learning approaches for automated crop disease detection. Curr. Plant Biol. **40**, 100382 (2024). https://doi.org/10.1016/j.cpb.2024.100382
7. H. Sun, R. Fu, X. Wang, Y. Wu, M.A. Al-Absi, Z. Cheng, Q. Chen, Y. Sun, Efficient deep learning-based tomato leaf disease detection through global and local feature fusion. BMC Plant Biol. **25**, 311 (2025). https://doi.org/10.1186/s12870-025-06247-w
8. S. Hemalatha, J.J.B. Jayachandran, A multitask learning-based vision transformer for plant disease localization and classification. Int. J. Comput. Intell. Syst. **17**, 188 (2024). https://doi.org/10.1007/s44196-024-00597-3
9. S.U. Khan, A. Alsuhaibani, A. Alabduljabbar, F. Almarshad, Y.N. Altherwy, T. Akram, A review on automated plant disease detection: Motivation, limitations, challenges, and recent advancements for future research. J. King Saud Univ. Comput. Inf. Sci. **37**, 34 (2025). https://doi.org/10.1007/s44443-025-00040-3
10. S. Kalaivani, C. Tharini, T.S. Viswa, K.F. Sara, S.T. Abinaya, ResNet-based classification for leaf disease detection. J. Inst. Eng. India Ser. B **106**, 1-

14 (2025). https://doi.org/10.1007/s40031-024-01062-7

11. P. Yadav, P. Singh, Disease detection techniques in plants: Transition from manual to automation, In Proceedings of International Workshop on New Approaches for Multidimensional Signal Processing, Springer, Singapore, December 03 (2022), 93-109
12. S. Duhan, P. Gulia, N.S. Gill, E. Narwal, RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification. Curr. Plant Biol. **42**, 100459 (2025). https://doi.org/10.1016/j.cpb.2025.100459
13. M. Rezaei, D. Diepeveen, H. Laga, M.G. Jones, F. Sohel, Plant disease recognition in a low data scenario using few-shot learning. Comput. Electron. Agric. **219**, 108812 (2024). https://doi.org/10.1016/j.compag.2024.108812
14. E. Moupojou, F. Retraint, H. Tapamo, M. Nkenlifack, C. Kacfah, A. Tagne, Segment anything model & fully convolutional data description for plant multi-disease detection on field images. IEEE Access **12**, 102592-102605 (2024). https://doi.org/10.1109/ACCESS.2024.3433495
15. X. Jiang, J. Wang, K. Xie, C. Cui, A. Du, X. Shi, W. Yang, R. Zhai, PlantCaFo: An efficient few-shot plant disease recognition method based on foundation models. Plant Phenomics **7**, 100024 (2025). https://doi.org/10.1016/j.plaphe.2025.100024
16. Z. Salman, A. Muhammad, D. Han, Plant disease classification in the wild using vision transformers and mixture of experts. Front. Plant Sci. **16**, 1522985 (2025). https://doi.org/10.3389/fpls.2025.1522985