Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Edge AI Framework for Early Groundnut Leaf Spot Diagnosis Using Lightweight Vision Transformers

Parthasarathy Seethapathy
Amrita Vishwa Vidyapeetham
Kannan M
Amrita Vishwa Vidyapeetham
Manivasagam V.S.
Amrita Vishwa Vidyapeetham
Gopakumar G
gopakumarg@am.amrita.edu

Amrita Vishwa Vidyapeetham

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11462499/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Edge AI Framework for Early Groundnut Leaf Spot Diagnosis Using Lightweight Vision Transformers

Parthasarathy Seethapathy[1,2*], Kannan M[1], V.S. Manivasagam[1], G Gopakumar[3*]

[1]Amrita School of Agricultural Sciences, Amrita Vishwa Vidyapeetham, Arasampalayam, Coimbatore, 642 109, Tamil Nadu, India.
[2]Current Affiliation: : Agrotechnology Division, CSIR - Institute of Himalayan Bioresource Technology, Palampur, 176061, Himachal Pradesh, India.
[3*]Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, 690525, Kerala, India.

*Corresponding author(s). E-mail(s): s_parthasarathy@cb.amrita.edu, parthasarathy.91@csir.res.in; gopakumarg@am.amrita.edu;
Contributing authors: m_kannan1@cb.amrita.edu; vs_manivasagam@cb.amrita.edu;

Abstract

Plant disease detection is essential for improving crop productivity, ensuring food security, and supporting sustainable agricultural practices. This study presents an edge AI framework using lightweight vision transformer for early and stage-specific groundnut leaf spot disease detection. An Attention-Augmented Lightweight Mobile Vision Transformer (AAL-MobileViT) framework is proposed by integrating spatial and channel attention mechanisms with the MobileViT backbone to improve lesion-focused feature representation and classification performance. The proposed framework was evaluated against multiple lightweight convolutional neural network (CNN) and base-line MobileViT models. A total of 6,210 real-time groundnut leaf images belonging to three classes, namely early leaf spot, healthy, and late leaf spot, were utilized for model training and evaluation. The experimental results confirmed that the proposed AAL-MobileViT model performs better than the other models by 94.88% classification accuracy on the test set. In addition, several experiments using seeds and an out-of-distribution

(OOD) test proved stable learning behavior and increased stability to an external field dataset. For practical agricultural use, a real-time mobile application called "Amrita Crop Care" was developed, to enable the diagnosis of diseases using the smartphone and give appropriate recommendations. The process of statistical verification, cross-location validation, and smartphone deployment exhibits the practical usefulness of the proposed framework for field-level groundnut leaf spot diagnosis.

Keywords: Plant disease detection, Groundnut leaf spot, CNN, AAL-MobileViT, Mobile applications, Smart farming systems

1 Introduction

Groundnut, or peanut is an herbaceous plant belonging to the family Fabaceae. It is an important crop for food security because its edible kernels and oil are widely used for food and commercial purposes (Tan et al. 2020). However, groundnut production is affected by several fungal diseases, including early leaf spot, late leaf spot, rust, stem rot, and collar rot. Among these, early leaf spot is one of the most important diseases because it occurs early during the crop cycle, rapid spread under humid conditions, and significant reduction in photosynthetic activity, which causes substantial yield loss (Hossain et al. 2024). In India, early leaf spot can cause a 30 to 50% reduction in pod yield under unmanaged conditions and may result in losses of up to 54% in severe endemic situations (Gupta 2022). Early and accurate diagnosis of leaf spot disease is therefore essential for protecting plant health and maintaining crop productivity (Buja et al. 2021). Early leaf spot generally appears 30–40 days after sowing, initially affecting older leaves before spreading rapidly under favourable environmental conditions. Late leaf spot usually develops later, around 50–60 days after sowing, and produces darker lesions that progress more aggressively (Thomidis et al. 2023). Both diseases intensify during the 60–100-day period, particularly during the post-flowering stage, and severe infections may cause extensive defoliation and substantial yield loss after 90–120 days (Buja et al. 2021 and Thomidis et al. 2023). Early-stage plant disease detection and appropriate disease management can decrease defoliation and reduce yield losses in cultivation (Hasanaliyeva et al. 2022). In field conditions, stage-wise leaf spot disease detection remains challenging due to symptoms often being similar to those of other foliar diseases (Barbedo 2019). Existing disease detection systems depend on manual field inspection by agricultural experts and farmers. This process is time-consuming and needs expert knowledge (Shoaib et al. 2023). The development and usage of smartphones gives support for real-time disease detection. It aids farmers in getting suitable information and disease management recommendations. To address the limitations of manual diagnosis, the proposed research focuses on real-time evaluation of early leaf spot symptoms in groundnut crops using image-based analysis (Asif et al. 2025). Deep learning (DL)-based computer-aided diagnosis systems have shown considerable potential for automated plant disease recognition in laboratory and controlled environments (Khan et al. 2021 and Qadri

et al. 2025). Convolutional Neural Network (CNN) models such as MobileNet, ShuffleNet, and EfficientNet-Lite have gained great attention because of their reduced processing complexity and adaptability for mobile deployment. However, CNN-based models primarily rely on local receptive field learning and often struggle to capture long-range contextual relationships and heterogeneous lesion patterns under complex field environments. Lightweight vision transformer (ViT)-based architectures such as MobileViT combine convolutional operations with transformer-based global attention and have shown performance for plant disease diagnosis. These architectures improve contextual feature representation and better discrimination of disease symptoms under varying agricultural conditions. However, only a limited number of studies have systematically compared lightweight CNN and lightweight transformer architectures for field-level plant disease diagnosis and mobile deployment. In addition, most existing studies focused on classification accuracy without inclusive assessment of practical deployment.

To address these research gaps, this study evaluates multiple lightweight CNN and transformer-based architectures for stage-specific classification of groundnut leaf spot disease and develops a smartphone-assisted disease diagnosis framework suitable for field deployment. In particular, an attention-improved MobileViT framework is proposed to improve lesion-focused feature learning through channel and spatial attention refinement while maintaining computational efficiency for mobile deployment. In addition, a real-time mobile application named "Amrita Crop Care" was developed with integrated cloud connectivity to support field-level disease monitoring and decision support.

The main contributions of this study include:

- **Field-acquired dataset collection:** A field-acquired dataset of groundnut leaf images was acquired and preprocessed under both controlled and field conditions, including healthy leaves and stage-wise leaf spot disease symptoms. A total of 6,210 plant-level images were collected using mobile cameras under varying lighting and climatic conditions to improve model generalization under field conditions during field-level deployment.
- **Proposed lightweight transformer framework:** A lightweight MobileViT-based framework, namely AAL-MobileViT, was developed for stage-specific groundnut leaf spot disease classification by integrating lightweight channel and spatial attention mechanisms with the MobileViT architecture.
- **Comparative evaluation of lightweight architectures:** Nine lightweight mobile-deployable CNN models and transformer-based architectures were evaluated to compare their classification performance, and suitability for real-time field-level deployment.
- **Explainability analysis using Grad-CAM:** Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the discriminative disease regions influencing model predictions, thereby improving model interpretability of the proposed disease diagnosis framework.

- **Computational efficiency analysis:** Model size, inference time, memory consumption, and power consumption were evaluated across multiple runs using different mobile devices to assess the computational efficiency and practical deployment potential for smartphone deployment of the proposed framework.
- **Statistical analysis and cross-location validation:** Cross-location external validation was conducted to evaluate the consistency and generalization performance of the proposed framework under unseen field conditions. In addition, statistical significance analysis using a paired t-test and p-value evaluation was performed to assess performance differences by the proposed model over baseline architectures.
- **Mobile application development:** A mobile application named "Amrita Crop Care" was developed using the Flutter framework to enable real-time field-level groundnut leaf spot disease detection and stage-wise classification and also to provide disease-stage-specific management recommendations.
- **Cloud-assisted disease monitoring:** Prediction results generated from the mobile application were integrated with cloud storage to support field-level disease monitoring and data management.

The remainder of this paper is organized as follows: Section II reviews the related work, Section III describes the proposed methodology and implementation, Section IV presents the experimental results and discussion, and Section V concludes the study with future directions.

2 Related work

This section reviews recent studies in CNN-based disease detection, attention-improved architectures, vision transformer approaches, and mobile deployment systems and highlights the research gaps relevant to this study. Table 1 lists the comparison of lightweight CNN and vision transformer architectures for mobile plant disease detection.

2.1 CNN-Based Plant Disease Detection

Thanjaivadivel et al. (2025) proposed an improved convolutional neural network that employs depth-wise separable convolution and inverted residual blocks to spot leaf diseases in plants. The findings of the trials also establish that the augmented convolutional neural network, outperforms other machine learning and deep learning models in classifying plant leaf diseases. The proposed model was not tested on a real-time dataset and may not be suitable for drone photographs. Anbumozhi et al. (2024) developed the Groundnut Leaf Diseases Identification and Classification using Convolutional Neural Network (GLDICCNN) model for multiclass classification of groundnut leaf diseases. The proposed model achieved a validation accuracy of 97.06%, and a test accuracy of 96%, showing higher classification performance than the compared baseline models, including ResNet50, ResNet101, and ResNet152. Balafas et al. (2023) evaluated object detection and classification methods using the PlantDoc dataset. YOLOv5 achieved higher disease localization ability, while ResNet50 and MobileNetV2 obtained classification accuracies of 61.01% and 59.74%, respectively. However, the

study reported limited stability under complex field conditions. Dheeraj and Chand (2024) developed a lightweight deep convolutional neural network named lightweight DenseNet (LWDN) for detection of plant leaf disease for agricultural applications. The suggested model has 93% fewer parameters than InceptionV3 and Xception and 90% and 50% fewer parameters compared to VGG16 and MobileNetV2. The limitation of this study includes, model deployment was not carried out in this study. Reddy et al. (2019) used image processing techniques, feature extraction and Support Vector Machine (SVM) classification for plant disease detection with k-means clustering. The proposed method achieved better early disease detection but was impacted by the noise in images, dependency of clusters, and limited field adaptability. Aishwarya et al. (2023) proposed a lightweight custom CNN (CCNN) for tomato leaf disease classification with 10 classes of diseases. The model showed comparable classification performance with very few hyperparameters and suggesting its suitability in mobile deployment applications. Seethapathy et al. (2025) compared various pre-trained CNN models for the detection of Dolichos yellow mosaic virus in lablab-bean leaves. The study, however, showed that larger and more diverse field datasets were required to better generalize. Vallabhajosyula et al. (2022) proposed a comprehensive taxonomy on the presentation of different pre-trained neural networks and presents the performance of a weighted ensemble of those models applicable to plant leaf disease detection. The performance of proposed model outperform state-of-the-art pre-trained models such as ResNet 50 and 101, InceptionV3, DenseNet 121 and 201, MobileNetV3, and NasNet. The limitation of this study includes no model deployment and used only open-source dataset, such as the Plant Village dataset. Pushpa et al. (2024) proposed CNN architectures for the detection of corn leaf diseases using extensive image preprocessing. The DenseNet-121 had an effective accuracy of 98.47% while ResNetV2 and EfficientNetV2B3 had an accuracy of 98.42% and 98.19% respectively. The models showed slight overfitting in the models, as evidenced by a difference between the training and validation accuracy. G. S. Damaraju et al. (2023) introduced an integrated deep learning system to detect early leaf spot disease in groundnut crops. Utilizing a custom CNN and ResNet-UNet, the model achieved 98.32% testing accuracy and precise severity estimations. However, limitations include high computational resource requirements, reliance on extensive annotated datasets, and a location-specific focus.

2.2 Attention-Based Disease Detection Methods

Attention mechanisms have recently gained considerable importance in plant disease detection by improving feature refinement and lesion localization ability (Remya et al. 2026). Xu et al. (2023) proposed an improved Xception-based architecture integrated with attention-augmented branches and a parts-activated feature fusion module for detecting peanut leaf diseases. The proposed model achieved improved classification accuracy, significantly outperforming baseline CNN architectures. Guo et al. (2024) developed a CNN-LSTM hybrid architecture integrated with Squeeze-and-Excitation (SE) attention mechanisms for peanut leaf spot disease prediction using meteorological and disease survey data. The integrated attention-based framework achieved higher prediction performance with $R^2 = 0.951$ and RMSE = 0.063, outperforming standalone CNN and LSTM models. Nigar et al. (2024) utilized

EfficientNetB0 integrated with explainable artificial intelligence (XAI) techniques such as LIME for multiclass plant disease classification across 38 plant diseases using 87,000 images. Although attention-enhanced CNN architectures significantly improve accuracy, lesion-focused feature extraction and disease localization capability, most existing methods remain computationally expensive and less appropriate for resource-constrained mobile deployment environments.

2.3 Vision Transformer-Based Agricultural Disease Detection

Hassija et al. (2025) reported that Vision Transformer (ViT)-based architectures have recently emerged as alternatives to conventional CNN models due to their ability to capture global contextual information through self-attention mechanisms. Unlike CNN architectures that primarily focus on local receptive fields, transformer-based models effectively learn long-range spatial dependencies and heterogeneous lesion patterns, which are particularly important under complex agricultural imaging conditions. Tianfang Zhang et al. (2026) reported that MobileViT integrates lightweight convolutional operations with transformer-based global attention mechanisms to achieve improved contextual feature representation while maintaining computational efficiency for edge devices. Transformer-based architectures have demonstrated higher stability under varying illumination conditions, complex backgrounds, and irregular disease symptom distributions. Shyam Singh Rajput et al. (2026) highlighted that only limited studies have explored lightweight transformer-based architectures for real-time agricultural disease diagnosis and smartphone-assisted field deployment. Furthermore, the integration of lightweight attention refinement mechanisms with transformer-based architectures remains relatively unexplored for detecting plant diseases in the field. Remya et al. (2024) proposed an integrated approach utilising vision transformers and acoustic sensors to detect whitefly infestations in cotton crops. It performed better than the benchmark architectures and reached high accuracy within 20 epochs. There are some restrictions such as need for extensive real world field tests to determine effectiveness.

2.4 Mobile and Edge Deployment for Plant Disease Diagnosis

For mobile deployment, lightweight architectures like MobileNet, ShuffleNet and EfficientNet-Lite have attracted interest for their small size, inference latency and suitability for mobile deployment Hu et al. (2026). Dat Ngo Yu Hu et al. (2025) presented that lightweight CNN models offer advantages in terms of inference time and computational efficiency, but they may exhibit limitations in terms of their ability to classify accurately under field agricultural settings with background noise, light variations, and varying lesion patterns. Liya Hu et al. (2026) reported that most existing studies primarily emphasize classification accuracy while providing limited analysis of deployment-oriented metrics such as inference time, memory consumption, power efficiency, explainability, out-of-distribution, and real-time cloud integration ability. These limitations restrict the practical use of existing systems for field-level agricultural deployment.

Table 1 Comparison of lightweight CNN and vision transformer architectures for mobile plant disease detection

Architecture	Attention / ViT mechanism	Mobile deployment	Key limitation
MobileViT / MobileViT V2	Hybrid CNN-transformer with global and separable self-attention	Optimized for smartphones, IoT, and low-power edge devices	Higher memory usage and inference latency than lightweight CNNs under resource-constrained deployment
Vision Transformer (ViT)	Global self-attention mechanism	Applicable to edge devices with optimized implementations	High computational complexity and limited performance on small-scale datasets
CAS-ViT	Convolutional additive self-attention with spatial and channel attention	Efficient deployment through CoreML and Apple Neural Engine	Performance scalability may decrease on very large-scale datasets
LeafB-Net	Dual-channel Efficient Channel Attention and Triplet Attention	Suitable for edge devices with low memory consumption and high frame rates	Sensitive to severe illumination variation and foliage occlusion
SwiftFormer	Efficient additive attention mechanism	Designed for real-time mobile vision applications	Attention operations are simplified to maintain low latency
EfficientFormer / EfficientFormerV2	Lightweight transformer with efficient self-attention refinement	Optimized for mobile inference with speed comparable to MobileNet	Performance may degrade under extremely small parameter configurations
Swin Transformer	Shifted window self-attention with hierarchical representation	Supports variable-resolution deployment	Complex window operations increase deployment complexity
MobileNet Series (V1/V2/V3)	Depthwise separable convolution with lightweight attention variants	Widely adopted for smartphones and embedded AI applications	Limited capability in capturing global contextual lesion information

2.5 Research Gap and Motivation

Although significant progress in DL-based plant disease detection, several limitations remain. Most existing studies primarily focus on conventional CNN architectures such as ResNet, EfficientNet, and MobileNet variants, while only limited research has explored lightweight transformer-based architectures for agricultural disease diagnosis. Furthermore, few studies have systematically compared lightweight CNN and lightweight vision transformer architectures under field deployment conditions. Existing approaches also lack field-collected data, plant-level splitting, cross-location validation, statistical testing, and mobile deployment. In addition, lightweight attention-improvedtransformer architectures for real-time smartphone-assisted agricultural applications remain largely less explored. To address these limitations, this study proposes the AAL-MobileViT frame work for accurate and computationally efficient classification of groundnut leaf spot disease. The proposed framework combines lightweight transformer-based contextual learning with channel and spatial attention refinement to improve lesion-focused feature representation while maintaining

deployment efficiency appropriate for smartphone-assisted applications for diagnosing diseases in the field.

3 Proposed methodology

The proposed research architecture given in Fig. 1 comprises three primary components: (i) field experiment, field-acquired dataset collection, and pre-processing (ii) disease prediction in groundnut plant leaves using different DL models and the proposed AAL-MobileViT framework; and (iii) mobile application and cloud integration.

3.1 Field Experiment

The groundnut field used for image acquisition in this study is given in Fig.2.

3.2 Field-acquired dataset collection

The field-acquired groundnut leaf dataset consists of 6,210 high-resolution images (3120 $\times$ 4160 pixels) of groundnut leaves. Images were captured using a Realme C2 mobile phone, equipped with a dual rear camera (13 MP primary sensor + 2 MP depth sensor) under both controlled (polyhouse) and open field conditions to capture diverse field and imaging conditions under standard agronomic management practices. Images were captured at close to the leaf to overcome motion blur, severe occlusion and duplicate or poor-quality images were removed. During dataset collection disease symptoms were independently assessed by two experts from plant pathologists with experience in groundnut disease diagnosis. The experts classified each image based on visible lesion characteristics, lesion distribution. Images with uncertain or ambiguous symptoms were reviewed by both experts, and a consensus label was assigned after discussion. The dataset includes three classes: early leaf spot, healthy, and late leaf spot. To avoid data leakage and inter-sample dependency, the dataset was partitioned at the plant, location, date, and acquisition-session levels. The training dataset was collected from 250 plants, with approximately 20 images captured per plant in November 2025 across ten acquisition sessions from the Amrita School of Agricultural Sciences, Coimbatore, India. The validation dataset was collected from 30 plants with approximately 20-21 images captured per plant in November 2025 across three acquisition sessions at Arasampalayam village, Coimbatore, India. The testing dataset was collected from 30 plants in November 2025, with approximately 20-21 images captured per plant across three acquisition sessions in Myleripalayam village, Coimbatore, India. All images from the same plant and acquisition session were assigned to only one subset to prevent overlap between training, validation, and testing datasets. The original dataset images are shown in Fig. 3.

3.3 Image preprocessing

The preprocessing stage helps to convert the acquired images into a suitable format for model training and evaluation with camera-captured images. It includes resizing, annotation, augmentation, and normalization. In addition to deploying the best model

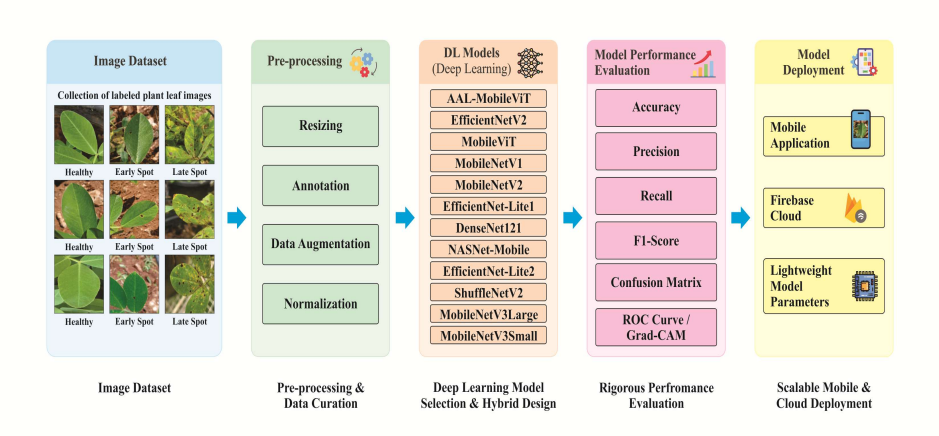

Fig. 1 The proposed framework of the deep learning-based disease classification system.



Fig. 2 Groundnut field used for image acquisition in this study.

on a resource-constrained mobile application, input image dimensions were reduced to reduce computational cost and reduce memory and processing requirements. For the dataset images, the size of the images was reduced from 3120×4160 to 224×224 to match the input size of the lightweight deep learning models (Mahmood et al. 2024). Let the original field-captured RGB groundnut leaf image be represented as a 3D tensor.

$$\mathbf{I} \in \mathbb{R}^{H_{\text{orig}} \times W_{\text{orig}} \times 3} \tag{1}$$

where $H_{\text{orig}} = 3120$ and $W_{\text{orig}} = 4160$ denote the original spatial height and width dimensions, respectively. The resized image tensor $\mathbf{I}_r \in \mathbb{R}^{224\times224\times3}$ is obtained by applying a spatial resizing operator $\mathcal{R}(\cdot)$ using bilinear interpolation to preserve high-frequency edge-boundary lesion features.

$$\mathbf{I}_r = \mathcal{R}\left(\mathbf{I}, 224, 224\right) \tag{2}$$

where $224 \times 224 \times 3$ represents the target spatial and channel dimensions required for subsequent feature extraction stages.

The pre-processed dataset was manually annotated by labelling relevant images with relevant metadata to form a structured dataset (Aishwarya and Reddy 2023). Two plant pathology experts independently annotated the images, and disagreements were resolved through joint review. The total number of images were then divided into 80% training, 10% validation, and 10% test images for further processing as shown in Table 2. To train the DL models, the training set was used, the validation set was used to check the performance of the DL models during the training process, and the test set was used to evaluate the final performance of the trained DL models (Vinay et al. 2025). Classes in the training set were imbalanced. In order to balance the dataset, the images of three classes were augmented by flipping, rotating, and shifting, with 2000 images per class in the training set. The augmentation techniques were selected as they were both simple and effective in diversifying the samples without changing the symptoms, as opposed to more complicated transformations (Cap et al. 2022). Moreover, MixUp augmentation was used during training to improve model generalization. In order to improve the stability and speed up the training of the model, the image was normalised by scaling the values of the pixels to the range (0,1) and dividing the image data by 255 to ensure that the input to the model remains consistent throughout the training, validation, and test sets (Dhingra et al. 2018).

Table 2 Dataset utilized in this study

Class	Training (80%)	Test (10%)	Validation (10%)
Early leaf spot	1400	175	175
Healthy leaves	2000	250	250
Late leaf spot	1600	180	180

3.4 Experimental protocol

The implementation and evaluation were performed on a workstation with an Intel Core i7-10700F processor, 16 GB of RAM, an NVIDIA GeForce GTX 1050 Ti GPU with 4 GB of memory, and a 64-bit Windows 11 Pro operating system. The experimental environment was well suited for model training, evaluation, and deployment analysis. To develop mobile application, android studio software used and Vivo T3 smartphone (6 GB RAM, Android 14) used as emulator.

3.5 CNN architectures

In this paper, several established CNN architectures were evaluated for lightweight plant disease classification, such as EfficientNetV2, MobileNetV1, MobileNetV2, MobileNetV3-Small, MobileNetV3-Large, ShuffleNetV2, DenseNet121, NASNet-Mobile, EfficientNet-Lite1, and EfficientNet-Lite2. These models were selected due to

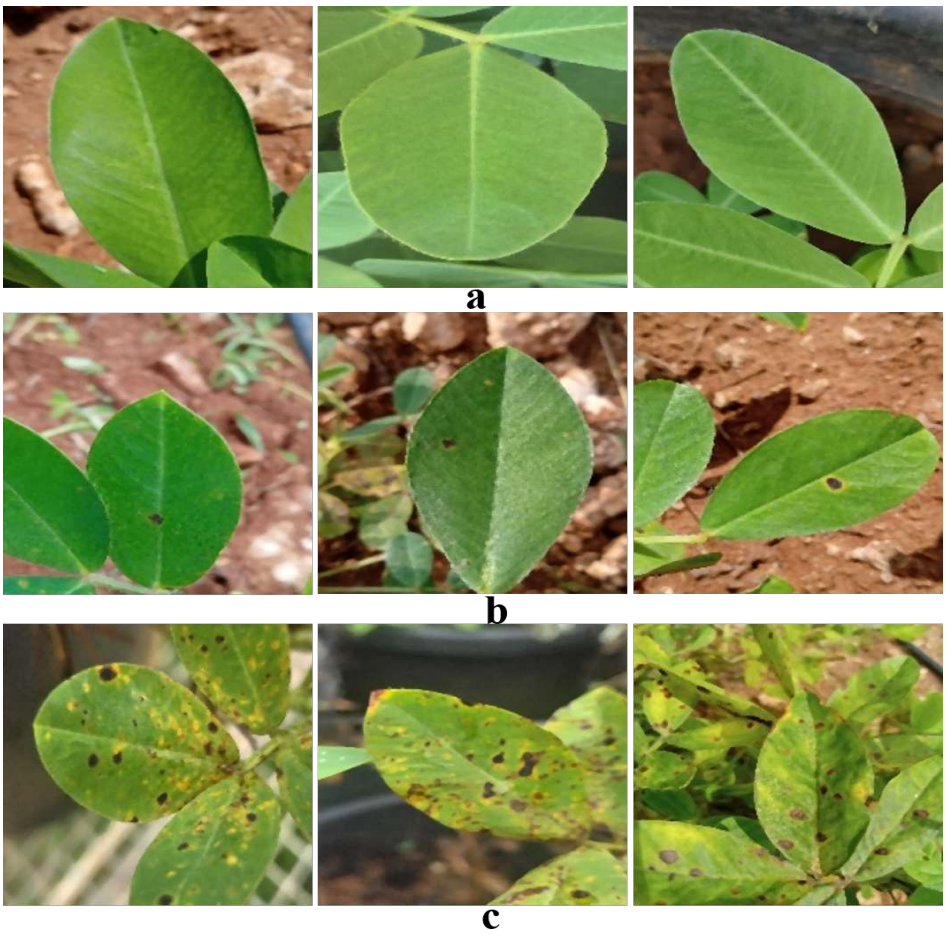

Fig. 3 Field-acquired dataset images (a) healthy leaves, (b) early leaf spot, and (c) late leaf spot.

their low computational cost, fewer model parameters, and suitability for smartphone deployment. EfficientNetV2 is a computationally efficient architecture widely used for plant disease classification. MobileNet-based architectures utilize depthwise separable convolutions to reduce computational cost while maintaining competitive classification performance. MobileNetV2 further introduces inverted residual bottleneck layers for improved feature extraction efficiency, whereas MobileNetV3 incorporates neural architecture search and lightweight attention mechanisms for optimized mobile inference. ShuffleNetV2 employs channel shuffling operations to improve feature communication with reduced latency, making it for resource-constrained devices.

3.6 Mobilevit Architecture

The MobileViT allows the architecture to preserve lightweight computational characteristics while improving feature representation compared with CNN models. In this

work, the MobileViT-XXS variant pretrained on the ImageNet dataset was adopted due to its reduced model complexity, mobile deployment suitability, and better classification performance for agricultural image analysis. All MobileViT experiments were conducted under the same preprocessing, augmentation, and experimental protocol settings used for the lightweight CNN benchmark models to enable a consistent comparison.

3.7 Proposed AAL-MobileViT

To improve lesion-focused feature learning and enhance classification stability, a lightweight attention mechanism was integrated with the baseline MobileViT frame work. The proposed AAL-MobileViT model combines the global contextual representation abilityof MobileViT with lightweight channel and spatial attention modules for disease feature extraction. To perform intermediate, multi-scale feature recalibration, the proposed framework integrates custom Attention-Augmented Lightweight (AAL) modules directly downstream of specific representation blocks. As illustrated in the structural pipeline in Fig. 4, instead of utilizing a single attention block at the terminal layer, we insert our AAL modules immediately after MobileViT Block 1 and MobileViT Block 2. These diagnostic patterns are frequently lost or blurred in standard deep learning networks during subsequent aggressive pooling operations. Each integrated AAL block sequentially applies channel-wise semantic scaling and spatial-localizing refinement to focus the feature maps, which are then reinforced with a residual skip connection to maintain stable gradient flow. The spatial attention module highlights lesion regions and improves disease localization under varying field conditions. The refined spatial features are further combined with the original backbone features using a residual skip connection to preserve important feature information and improve training stability. Finally, the refined feature maps are passed through a classification head consisting of global average pooling, dropout regularization, a fully connected dense layer with 256 neurons, and a softmax output layer for three-class disease classification. Table 3 shows the hyperparameters of the proposed model.

The intermediate feature map extracted by the MobileViT core is represented as a 3D tensor $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$. The proposed Attention-Augmented Lightweight (AAL) module processes this tensor sequentially to recalibrate channel-wise and spatial-localizing feature maps.

First, the channel-recalibrated feature map $\mathbf{F}_c \in \mathbb{R}^{H \times W \times C}$ is obtained by scaling the input feature map element-wise with the adaptive channel attention weights:

$$\mathbf{F}_c = \mathbf{F} \otimes \sigma\left(\mathbf{W}_2 \cdot \delta\left(\mathbf{W}_1 \cdot \mathrm{GAP}(\mathbf{F})\right)\right) \tag{3}$$

where $\mathrm{GAP}(\cdot)$ represents the global average pooling operator, $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ represent the trainable weights of the bottleneck fully connected layers with a reduction ratio $r = 16$, $\delta(\cdot)$ denotes the rectified linear unit (ReLU) activation function, $\sigma(\cdot)$ represents the sigmoid activation function, and $\otimes$ denotes element-wise multiplication.

Second, the spatial attention map $\mathbf{M}_s \in \mathbb{R}^{H \times W \times 1}$ is generated to geometrically localize irregular foliar lesion contours by aggregating spatial statistics along the

channel dimension:

$$\mathbf{M}_s = \sigma\left(f^{7\times7}\left([\text{AvgPool}_c(\mathbf{F}_c); \text{MaxPool}_c(\mathbf{F}_c)]\right)\right) \tag{4}$$

where $\text{AvgPool}_c(\cdot)$ and $\text{MaxPool}_c(\cdot)$ denote channel-wise average and maximum pooling operations, respectively, $[\cdot;\cdot]$ represents channel-wise tensor concatenation, and $f^{7\times7}$ denotes a standard 2D spatial convolution operation with a 7×7 filter.

The spatially refined feature representation $\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$ is subsequently formulated as:

$$\mathbf{F}_s = \mathbf{F}_c \otimes \mathbf{M}_s \tag{5}$$

where $\otimes$ represents the element-wise multiplication operator.

To prevent gradient vanishing and preserve high-frequency texture details from the earlier layers, a residual skip connection is integrated to produce the final refined feature representation $\mathbf{F}_r \in \mathbb{R}^{H\times W\times C}$:

$$\mathbf{F}_r = \mathbf{F} \oplus \mathbf{F}_s \tag{6}$$

where $\oplus$ represents the element-wise addition operator.

Finally, the target probability distribution over the three target pathological classes ($K = 3$: healthy, early leaf spot, and late leaf spot) is computed by projecting the final global pooled vector through a softmax classification head:

$$P(y_i) = \frac{\exp(z_i)}{\sum_{j=1}^{3}\exp(z_j)} \tag{7}$$

where z_i denotes the output logit for class i, and $P(y_i)$ represents the predicted probability of the i-th crop class.

Table 3 Hyperparameters of the proposed AAL-MobileViT model

Hyperparameter	**Value**
Input size	224×224
Batch size	32
Loss function	Categorical cross-entropy
Activation function	ReLU (dense layer), Softmax (output layer)
Optimizer	Adam
Learning rate	0.0001
Epochs	25
Dropout rate	0.5
Training/validation/test split	80% / 10% / 10%

3.8 Mobile Application and Cloud Integration

A mobile application was developed using the Flutter framework and Dart programming language for real-time groundnut leaf spot disease detection in the field. The

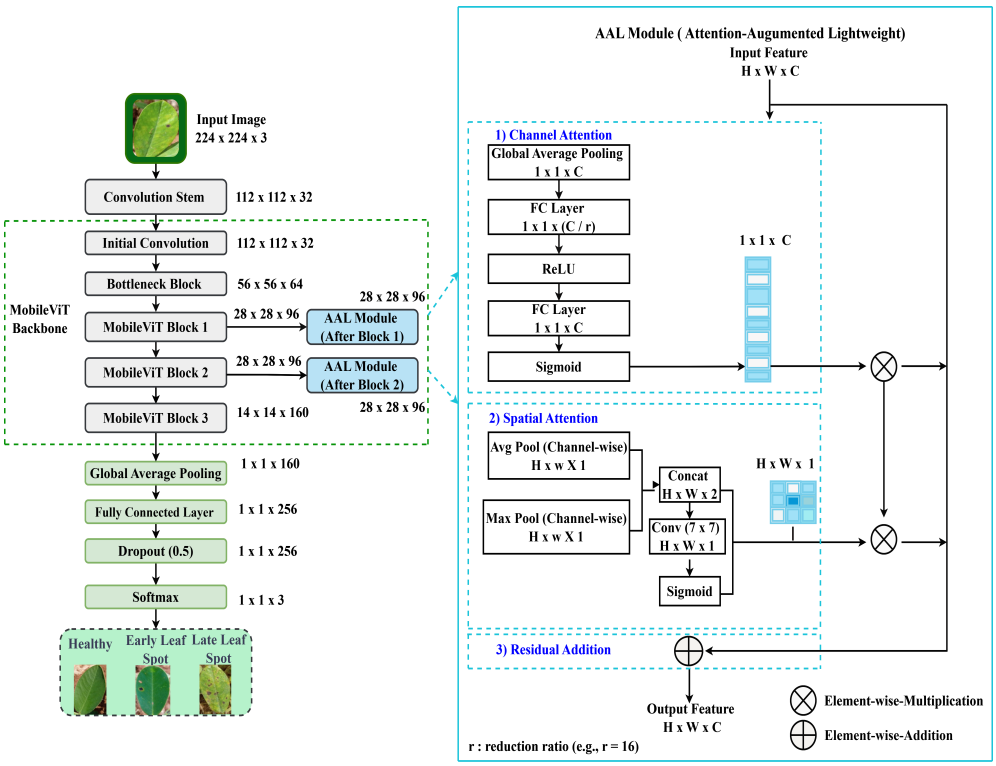

Fig. 4 Proposed AAL-MobileViT model with the placement of attention modules.

trained deep-learning model was integrated into the application using TensorFlow Lite for on-device inference. Initially, the trained `.h5` model was converted into the lightweight `.tflite` format to reduce memory usage and computational overhead while improving inference speed on mobile devices. The developed application supports both gallery-based image classification and real-time image acquisition through the smartphone camera, allowing image classification from both the gallery and the smartphone camera, as illustrated in Fig. 5. Upon image acquisition, the integrated AAL-MobileViT model performs on-device inference and predicts the disease category among early leaf spot, healthy, and late leaf spot classes and provides disease management recommendations.

For real-time data management and disease monitoring, firebase cloud database integration was used as the cloud storage platform. For each prediction instance, the system records the predicted disease category, user-provided field location, unique user identification information, and coordinated universal time (UTC) timestamp. These stored records support disease tracking, field monitoring, and analysis of disease occurrence across locations across different agricultural locations.

4 Results and discussion

4.1 Classification Performance

Table 4 presents the comparative classification performance of the evaluated lightweight CNN and transformer-based architectures for groundnut leaf spot disease detection. The performance evaluation was conducted using accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC) metrics to fully assess the disease classification performance of each model under identical experimental settings. Among all evaluated architectures, the proposed AAL-MobileViT model achieved the highest classification performance with an accuracy of 94.88%, precision of 0.9493, recall of 0.9420, F1-score of 0.9444, and AUC of 0.9954. The results indicate that the integration of lightweight channel and spatial attention mechanisms improved lesion-focused feature representation and disease discrimination ability. The CNN-based EfficientNetV2 and baseline MobileViT with accuracies of 94.55% and 93.22%, respectively, performing better than most conventional lightweight CNN architectures. Unlike CNN architectures that primarily depend on local receptive fields, MobileViT-based architectures effectively captured long-range contextual relationships and heterogeneous lesion characteristics through self-attention mechanisms. This ability proved particularly beneficial for distinguishing visually complex disease symptoms under varying field conditions. Among the CNN models, MobileNetV1, MobileNetV2, EfficientNet-Lite1, and DenseNet121 achieved comparatively balanced classification performance with accuracies ranging from 89% to 90%.

In contrast, MobileNetV3-Small and MobileNetV3-Large exhibited substantially lower performance, suggesting limitations in representing complex disease symptom variations within the considered dataset. ShuffleNetV2, NASNet-Mobile, and EfficientNet-Lite2 also demonstrated moderate performance. The proposed framework loss and accuracy curve is given in Fig. 6. To avoid overfitting and minimize unnecessary computational overhead, an early stopping criterion was applied during training, terminating the optimization process at epoch 16 when the validation loss ceased to improve. The high AUC value achieved by the proposed framework confirms its good class discrimination across all disease categories. In addition, the confusion matrix and ROC curve analysis demonstrated improved class separability and reduced inter-class misclassification, particularly between early leaf spot and late leaf spot classes, which often exhibit visually similar lesion patterns and are given in Fig. 7 and Fig. 8, respectively. The per-class performance analysis of the proposed framework further showed high classification reliability across all target disease categories (Table 5). The healthy class achieved the highest performance with a precision of 0.9579, recall of 1.0000, and F1-score of 0.9785, indicating reliable healthy leaf identification. The early leaf spot class achieved an F1-score of 0.9162, while the late leaf spot class recorded a high F1-score of 0.9384. The proposed framework demonstrated balanced and reliable classification performance across all classes, with only marginal performance drops for late leaf spot due to its variable symptoms and overlapping lesion characteristics.

To evaluate the temporal early-detection performance of the proposed AAL-MobileViT framework, we analyzed model sensitivity relative to the chronological development of groundnut foliar pathogens. Under typical field conditions, early leaf

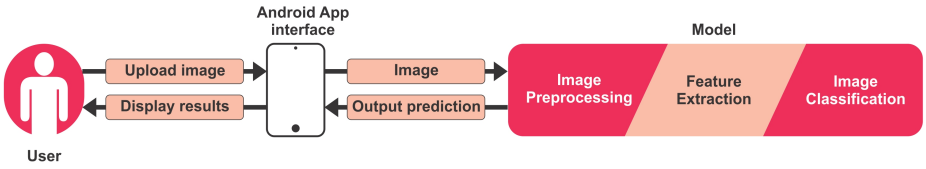

Fig. 5 Interface of android-based smartphone application for deployment

spot symptoms emerge on older lower leaves 30–40 days after sowing. Meanwhile, late leaf spot lesions appear during the crop's post-flowering stage at approximately 50–60 days after sowing. Our framework classifies both early leaf spot and late leaf spot infections at these critical initial symptomatic stages. This detection of early symptoms is mathematically enabled by our intermediate dual-attention modules (AAL), which prevent tiny necrotic spots and chlorotic yellow halos from being lost during progressive spatial downsampling. Catching these infections at this early temporal threshold allows farmers using the "Amrita Crop Care" application to apply targeted crop protection. They can address before the diseases spread aggressively during the 60–100 days after sowing post-flowering window. Overall, the experimental results validate that the proposed AAL-MobileViT framework maintains an effective trade-off between diagnostic accuracy, on-device deployment efficiency, and consistent lesion feature representation. The resulting edge-AI framework is suitable for real-time groundnut disease identification assisted by a smartphone directly in the field.

Table 4 Classification performance comparison of lightweight deep learning models

Model	Accuracy (%)	Precision	Recall	F1-score	AUC
AAL-MobileViT	**94.88**	**0.9493**	**0.9420**	**0.9444**	**0.9954**
EfficientNetV2	94.55	0.9413	0.9406	0.9394	0.9920
MobileViT	93.22	0.9336	0.9232	0.9265	0.9913
MobileNetV1	90.25	0.9094	0.8963	0.8970	0.9699
MobileNetV2	90.25	0.9063	0.8975	0.8965	0.9773
EfficientNet-Lite1	90.25	0.9029	0.8901	0.8929	0.9656
DenseNet121	89.75	0.9120	0.9007	0.8962	0.9751
NASNet-Mobile	89.09	0.8966	0.8797	0.8838	0.9665
EfficientNet-Lite2	88.60	0.8953	0.8727	0.8764	0.9717
ShuffleNetV2	86.28	0.8690	0.8695	0.8624	0.9557
MobileNetV3-Large	64.46	0.6677	0.6536	0.6393	0.7873
MobileNetV3-Small	55.70	0.5776	0.5710	0.5389	0.7311

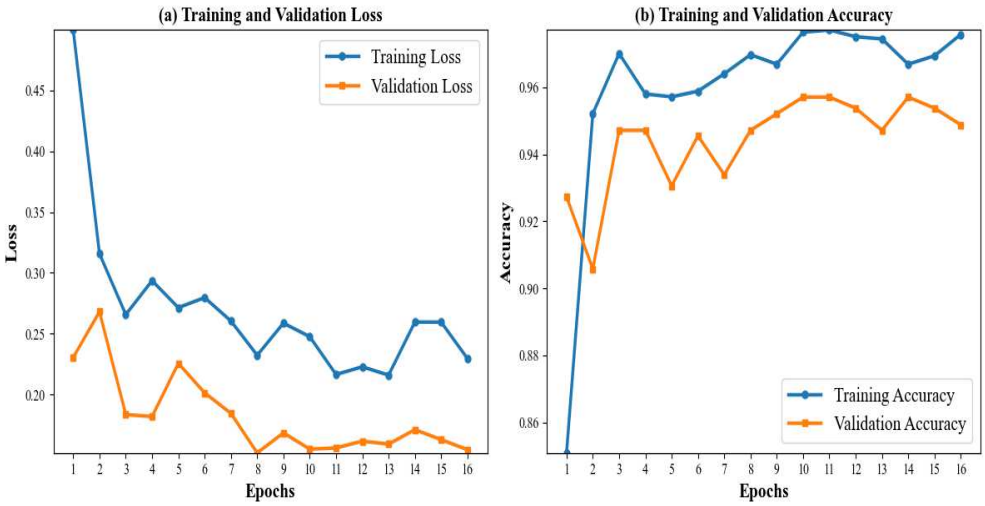

Fig. 6 Training and validation loss and accuracy curves of the proposed AAL-MobileViT model.

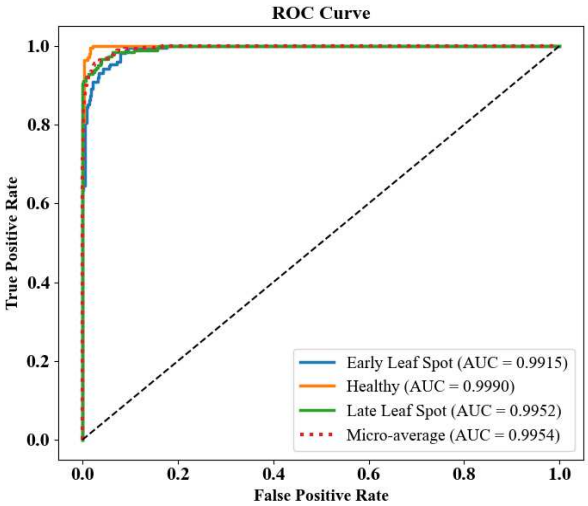

Fig. 7 Receiver operating characteristic (ROC) curve of the proposed AAL-MobileViT model.

4.2 Statistical Significance Analysis

To assess the stability, repeated experiments were conducted using different random seed values. The statistical stability analysis results for MobileViT and the proposed AAL-MobileViT framework across 15 randomized seeds are presented in Table 6. The proposed AAL-MobileViT framework showed consistent classification performance across all experimental runs, indicating similar performance across runs

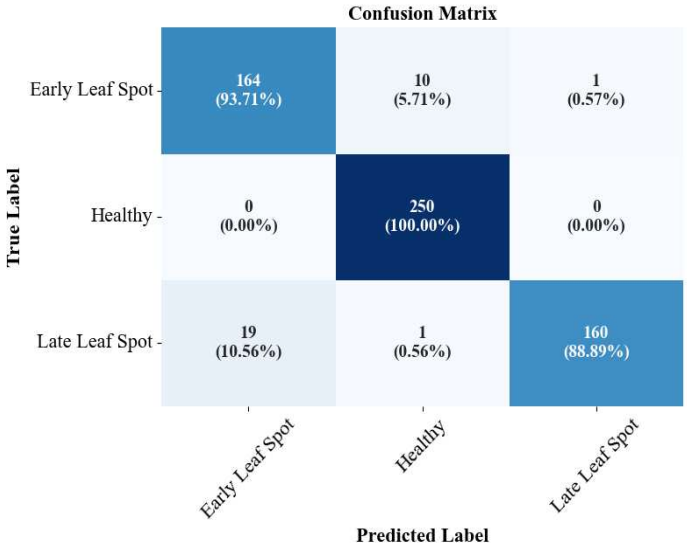

Fig. 8 Confusion matrix of the proposed AAL-MobileViT model for disease classification.

Table 5 Per-Class Performance Analysis of the Proposed Model

Class	Precision	Recall	F1-score
Early leaf spot	0.8962	0.9371	0.9162
Healthy	0.9579	1.0000	0.9785
Late leaf spot	0.9938	0.8889	0.9384

Table 6 Statistical stability of the proposed AAL-MobileViT and baseline MobileViT across 15 randomized seeds (mean ± standard deviation).

Model	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	AUC (%)
AAL-MobileViT	**94.66 ± 0.24**	**94.41 ± 0.24**	**94.42 ± 0.36**	**94.40 ± 0.33**	**98.71 ± 0.14**
MobileViT (Baseline)	94.18 ± 0.35	94.39 ± 0.26	94.20 ± 0.39	94.32 ± 0.46	98.65 ± 0.28

and consistent feature generalization performance. The model achieved a mean accuracy of $94.66 \pm 0.24\%$, precision of $94.41 \pm 0.24\%$, recall of $94.42 \pm 0.36\%$, F1-score of $94.40 \pm 0.33\%$, and AUC of $98.71 \pm 0.14\%$. Similarly, the baseline MobileViT architecture achieved a mean accuracy of $94.18 \pm 0.35\%$ and an AUC of $98.65 \pm 0.28\%$, showing similar classification performance across runs under repeated training conditions. To assess whether the difference in performance was statistically significant, a paired t-test was performed. As shown in Table 7, the proposed model achieved a

mean accuracy of 94.66% (95% CI: 94.53-94.79). The paired t-test yielded p = 0.031, indicating that the improvement over the baseline MobileViT model was statistically significant. Also, the effect size (Cohen's d = 0.63) indicates a moderate effect, demonstrating that the proposed attention-augmented architecture obtained higher mean performance than the baseline model across repeated experiments.

4.3 Ablation study of attention modules and Cross-Location External Validation

To investigate the contribution of the individual attention modules within the proposed framework, an ablation study was conducted by successively adding channel attention and spatial attention modules into the baseline MobileViT architecture. The comparative results are given in Fig. 9. The baseline MobileViT architecture achieved an accuracy of 93.22%. Incorporating only the channel attention module resulted in small improvement in precision and AUC; however, the overall classification accuracy slightly decreased to 93.06%. Similarly, the MobileViT model integrated with only spatial attention achieved an improved accuracy of 93.88% and demonstrated better lesion localization, indicating the effect of spatial feature refinement for disease symptom identification. The proposed AAL-MobileViT model, which combines both channel and spatial attention mechanisms, achieved the highest overall performance with an accuracy of 94.88%, precision of 0.9493, recall of 0.9420, F1-score of 0.9444, and AUC of 0.9954.

The obtained results show that the combination of channel-wise semantic refinement and spatial lesion localization improved feature extraction and classification under varying field conditions. To evaluate the generalization performance of the proposed framework under independent field conditions, a cross-location external validation was performed using an independent dataset comprising 500 field images collected from two villages, Kulathupalayam and Andipalayam, Coimbatore, India. during December 2025. Images were acquired using three different smartphone devices (Realme C2, Redmi Note 12, and Samsung Galaxy S23) across five independent acquisition sessions under varying illumination and environmental conditions. The cross-location external validation dataset consisted of 170 healthy, 170 early leaf spot, and 160 late leaf spot images.

All evaluated models exhibited a reduction in performance compared with the internal test dataset due to variations in illumination, background complexity, camera characteristics, and environmental conditions across different locations. However, this performance reduction (approximately 10%) reflects the challenges encountered in field field deployment and provides a additional assessment under field conditions of model generalization. The comparative cross-location external validation performance results are given in Fig. 10. Among the evaluated architectures, the proposed AAL-MobileViT achieved the highest cross-location external validation accuracy of 84.12% demonstrating higher classification performance on the external dataset. The EfficientNetV2 achieved an accuracy of 81.02%, while baseline MobileViT achieved 74.10%. The results show that the integrated channel and spatial attention mechanisms enhance the representation of the lesions and help the proposed framework to better handle variation in field images.

Table 7 Statistical significance analysis comparing AAL-MobileViT and MobileViT

Comparison	Statistical test	*P*-value	Mean accuracy (95% CI)	Cohen's *d*
AAL-MobileViT vs. MobileViT	Paired *t*-test	0.031	94.66% (94.53–94.79)	0.63

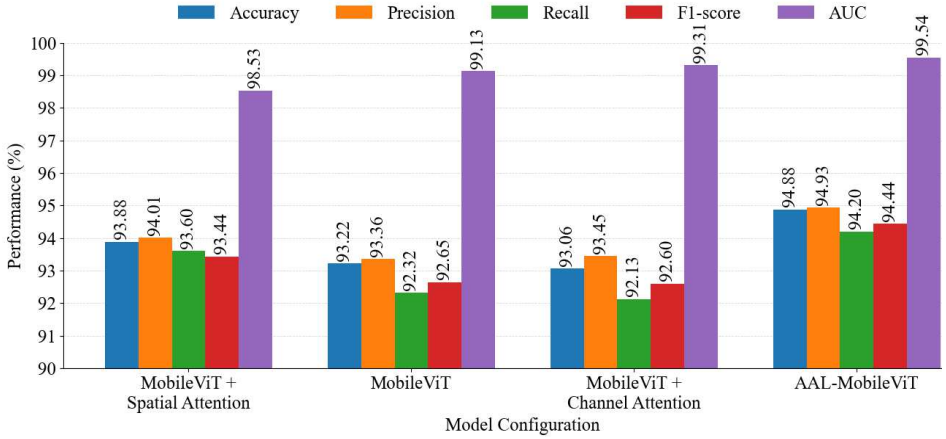

Fig. 9 Ablation study of attention modules in the MobileViT framework based on accuracy, precision, recall, F1-score, and AUC.

4.4 Computational Efficiency Analysis

For the evaluation of the practicality of the proposed framework for practical deployment, the computational efficiency analysis was carried out on five different kinds of smartphones. The assessment took into account model size, average inference time, average memory usage and average power usage in the real-time mobile deployment scenario. Each experiment was done with more than 100 inference runs and the mean ± standard deviation values were recorded. The comparative results are presented in Table 8. The proposed AAL-MobileViT framework achieved an average inference time of 138.42 ± 14.83 ms, outperforming both the baseline MobileViT and MobileNetV2 architectures. The reduced inference latency achieved the effectiveness of the lightweight attention integration strategy for real-time disease diagnosis on mobile devices with limited computational resources. It should be noted that the deployment performance of the proposed AAL-MobileViT is evaluated based on practical on-device performance metrics rather than model size alone. In particular, deployment efficiency refers to the model's ability to achieve low inference latency, moderate memory consumption, and reduced power consumption while maintaining high classification accuracy during real-time smartphone deployment. Although the proposed model incorporates lightweight attention modules and has a marginally larger memory

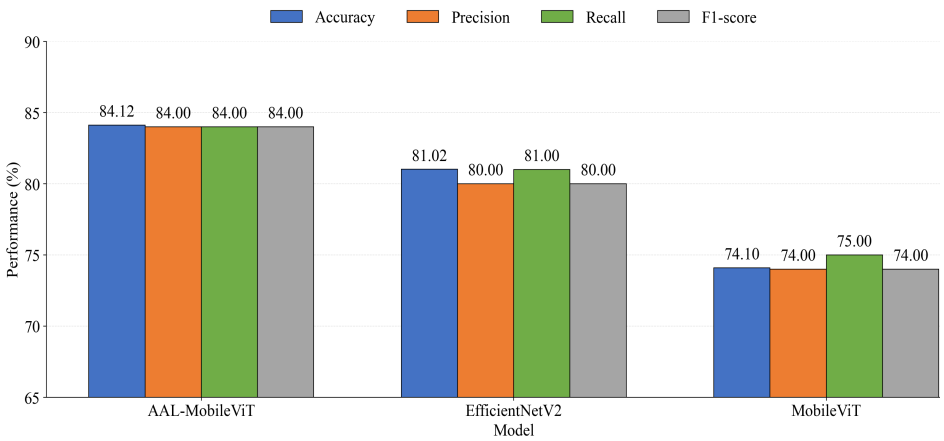

Fig. 10 Performance comparison of the proposed AAL-MobileViT and baseline models on the cross-location external validation dataset.

footprint than the baseline MobileViT, the experimental results demonstrate that it maintains faster response time and lower power consumption, making it appropriate for resource-constrained mobile agricultural applications.

Table 8 Mobile deployment efficiency across five smartphone devices

Model	Size (MB)	Time (ms)	Memory (MB)	Power (mW)
AAL-MobileViT	13.8	**138.42 ± 14.83**	**426.18 ± 17.25**	**2268.35 ± 34.72**
MobileNetV2	11.8	156.27 ± 16.91	440.73 ± 19.84	2294.18 ± 37.65
MobileViT	13.0	164.32 ± 18.45	418.85 ± 21.36	2386.42 ± 41.28

Although the proposed AAL-MobileViT incorporates additional channel and spatial attention modules, the computational overhead introduced by these modules is minimal because they operate on intermediate feature representations and require only lightweight feature recalibration. As a result, the backbone network remains the dominant contributor to the overall computational cost. The experimental results demonstrate that the additional attention operations do not introduce a noticeable increase in inference latency while improving feature representation and classification accuracy. Therefore, the proposed architecture maintains real-time inference performance and is applicable for deployment on resource-constrained mobile devices. In addition, the proposed framework exhibited moderate memory consumption of 426.18 ± 17.25 MB and lower average power consumption of 2268.35 ± 34.72 mW.

Although the proposed model requires a slightly larger memory footprint than MobileViT due to the inclusion of lightweight attention modules, the increase is marginal and does not directly determine inference speed. Although AAL-MobileViT has a slightly larger memory footprint than MobileViT, it showed lower inference

latency on the tested smartphones. This indicates that model size alone did not determine the measured inference time. The observed latency is also affected by the computational operations and their execution on the mobile hardware. The measured response time confirms that the proposed model maintains real-time performance despite the modest increase in memory usage. Although MobileNetV2 exhibited the smallest model size of 11.8 MB, its classification performance was substantially lower than the proposed transformer-based framework. In contrast, the proposed AAL-MobileViT achieved higher classification accuracy while maintaining computational efficiency applicable for real-time agricultural deployment. The baseline MobileViT architecture demonstrated comparatively higher inference time and power consumption, indicating that the proposed lightweight attention refinement improved both feature representation and deployment efficiency. Overall, the obtained results confirm that the proposed AAL-MobileViT provides an useful tradeoff between classification performance and computational efficiency.

4.5 Explainability

To improve the interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) analysis was performed to visualize the discriminative regions utilized during disease classification. The explainability results for both correctly and incorrectly classified samples are illustrated in Fig. 11. For correctly classified early leaf spot samples, the generated activation maps highly focused on lesion-affected regions containing characteristic necrotic spot patterns, indicating that the proposed AAL-MobileViT framework focused on disease-affected regions. Similarly, correctly classified late leaf spot samples exhibited concentrated activations around dense symptomatic lesion regions, demonstrating localization of lesion regions under field conditions. The explainability analysis further revealed that the integrated channel and spatial attention mechanisms focused more on lesion regions by suppressing background information and emphasizing disease-specific regions. The activation maps showed that the proposed framework primarily attended to visually significant symptomatic areas instead of irrelevant soil or surrounding vegetation regions. For incorrectly classified samples, Grad-CAM visualizations indicated that misclassification mainly occurred under challenging imaging conditions such as overlapping leaves, complex background interference, illumination variation, and visually similar lesion characteristics between early leaf spot and late leaf spot classes.

In certain cases, the attention maps partially focused on surrounding background regions, leading to reduced confidence scores and inter-class confusion. Overall, the explainability results demonstrate that the proposed AAL-MobileViT provides lesion-region localization and interpretable prediction patterns, thereby improving the transparency and reliability of smartphone-assisted groundnut disease diagnosis.

4.6 Mobile Deployment Results

The developed mobile application was evaluated under real-time field deployment conditions using five different smartphone devices to assess the practical use of the

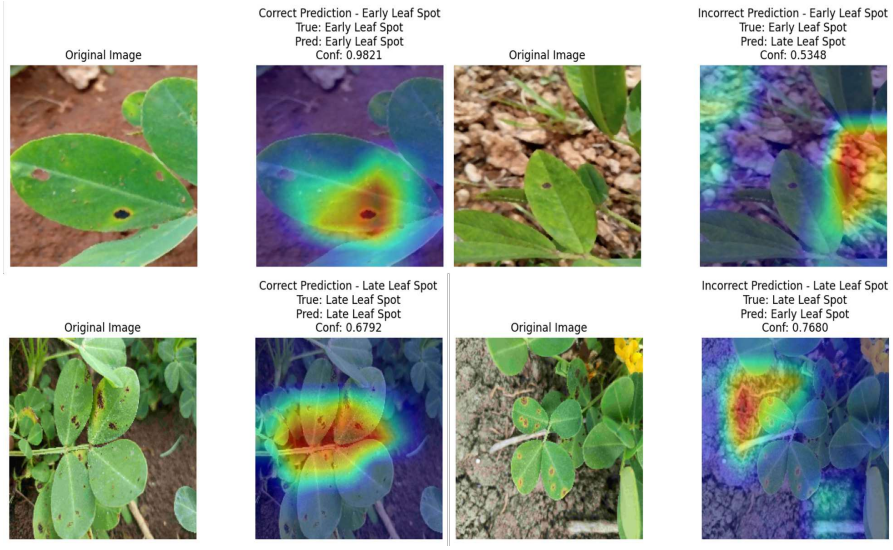

Fig. 11 Grad-CAM visualization of the proposed model for correct and incorrect prediction outputs.

proposed AAL-MobileViT. The application performed both gallery-based and real-time camera-based groundnut leaf disease classification with consistent inference time across varying environmental conditions. A confidence threshold of 0.75 was used as a functioning standard to reduce uncertain predictions in the mobile application. This threshold value was selected based on different test cases with mobile applications. Predictions with confidence greater than 0.75 are displayed as the detected disease, while predictions below 0.75 prompt the user to capture or upload a clearer leaf image. The application works so that the user provides an image containing the target plant leaf; no separate leaf/non-leaf validation module was implemented. Therefore, non-leaf or out-of-domain inputs are considered a limitation of the current system. In the future, the crop selection module will be attached to the mobile application for improved prediction.

The proposed framework demonstrated efficient on-device inference performance with lower inference latency, allowing near real-time classification that is appropriate for field-level agricultural applications. Sample outputs of the proposed "Amrita Crop Care" mobile application is given in Fig. 12 and Fig. 13. In addition, the lightweight architecture ensured stable application operation without noticeable delay in the application interface or the device heating during the continuous inference process. The prediction information, such as disease category, time, user ID, and field location information for each inference instance were successfully recorded in the firebase cloud. This allowed data to be managed centrally and spatiotemporally monitored the occurrence patterns of the disease in various agricultural areas. The developed system provides a mobile-based solution for low-cost and portable disease monitoring of groundnut diseases and the precision agriculture applications using the smartphone.

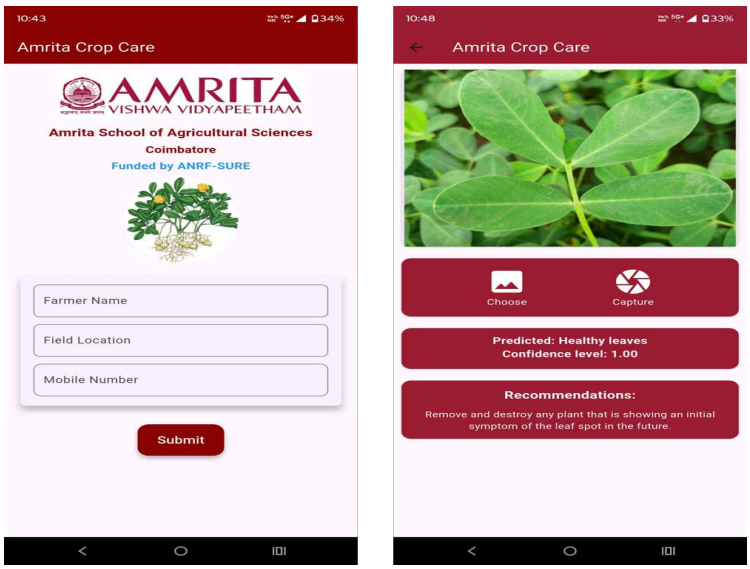

Fig. 12 Proposed "Amrita Crop Care" mobile application.

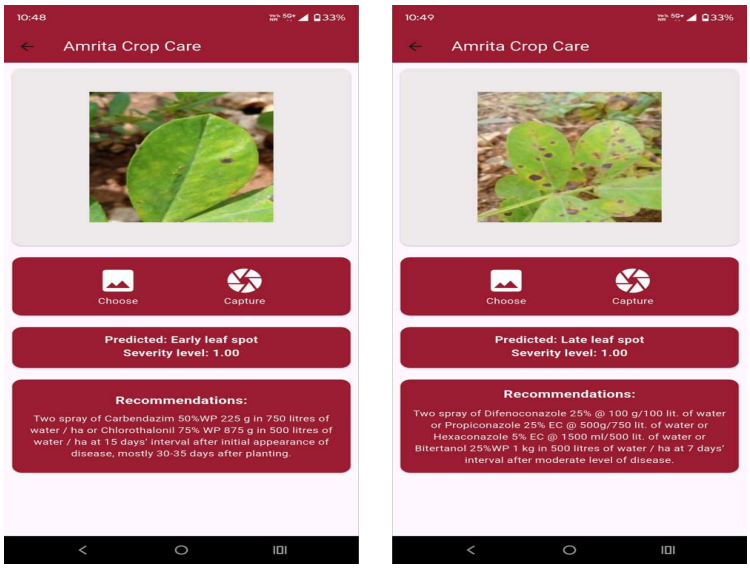

Fig. 13 User interface of the proposed "Amrita Crop Care" mobile application.

4.7 Discussion

The experimental results show that the proposed AALMobileViT achieved higher classification performance than the conventional lightweight CNN architectures for detecting the groundnut leaf spot disease. The higher performance may be related to the global self-attention mechanism of MobileViT that captures the long-range spatial

dependency and diverse lesion attributes in field-level agricultural images. In contrast to the typical CNN design, which focuses mainly on local receptive fields, MobileViT integrates the convolutional operations with the transformer-based contextual learning mechanism to simultaneously learn local texture information and global relationships among lesions. This performance is useful for distinguishing early leaf spot and late leaf spot symptoms, where lesion distribution, spatial arrangement, and contextual disease patterns play an important role in accurate classification. In addition, the transformer-based attention mechanism may improve performance under varying illumination conditions, complex backgrounds, and partial leaf occlusions commonly encountered in real agricultural environments.

The proposed AAL-MobileViT further improved classification performance through the integration of lightweight channel and spatial attention mechanisms. The channel attention module adaptively refined informative disease-related feature channels, while the spatial attention module improved lesion localization by emphasizing visually significant symptomatic regions. Grad-CAM explainability analysis confirmed that the proposed model focused on lesion-affected regions while reducing responses to background regions, thereby improving class discrimination. In addition to improving classification performance, the proposed AAL-MobileViT enhances the characterization of disease-specific visual patterns associated with different stages of groundnut leaf spot. The integrated channel and spatial attention mechanisms allow the framework to pay attention to the small differences of the lesions, their necrosis areas, and color variations, which are important for distinguishing between healthy leaves and early and late leaf spot symptoms. This may improve disease classification and diagnosis in the stage under a variety of field conditions. The MobileNet showed relatively low classification results due to its learning of local receptive fields. The obtained results show the trade-off between accuracy and deployment feasibility on a smartphone. The transformer-based architectures achieve high feature representation capabilities for accurate classification, and the drawback is higher computational complexity and latency of inference compared to lighter CNN models. The proposed AAL-MobileViT effectively provides a balance between accuracy and deployment requirements efficiency. The integration of lightweight attention refinement, transformer-based contextual learning, and deployment optimization resulted in a practical smartphone-assisted disease diagnosis framework for field precision agriculture applications. Although the proposed AAL-MobileViT demonstrated good classification performance for groundnut leaf spot classification, some limitations remain. The current study focuses on the classification of three disease categories using RGB images collected under specific field conditions. Future research should investigate multi-disease diagnosis, disease severity estimation, and the integration of multimodal data, such as hyperspectral images and environmental information, to further improve reliability and decision support. In addition, validating the proposed framework across diverse geographical regions, crop varieties, and growing seasons would further strengthen its generalization ability for practical precision agriculture applications.

5 Conclusion

This study presented a lightweight deep-learning framework for groundnut leaf spot disease classification using transformer-based architectures. The proposed AAL-MobileViT integrated MobileViT with lightweight channel and spatial attention mechanisms to improve lesion feature representation and disease classification. Using a field-acquired dataset consisting of 6,210 groundnut leaf images across three classes, namely early leaf spot, healthy, and late leaf spot, the proposed framework achieved a classification accuracy of 94.88%, outperforming several lightweight CNN architectures and baseline transformer models. The experimental results showed that the performance of recognizing the lesion pattern and complex field-level disease characteristics for heterogeneous patterns was improved using the transformer-based contextual learning compared to the conventional lightweight CNN architectures. Moreover, the proposed framework ensured low computational cost and suitability for mobile deployment.

In addition to improving classification accuracy, the proposed AAL-MobileViT can help to characterize groundnut leaf spot visual symptoms at various disease stages. The proposed framework focuses on lesion morphology, spatial distribution, and the developmental stages of the disease symptoms, which supports distinction between healthy leaves from early and late stages of disease in the field. This ability can be used to enable farmers to make disease management decisions, have better precision crop management. A mobile application, called "Amrita Crop Care" , was developed using the Flutter framework, and Firebase cloud services, for ease of use in the field. The developed application was used to provide real-time disease classification and recommendation, cloud-based data storage and spatiotemporal disease monitoring in field conditions. Although the proposed framework demonstrated high classification performance, the study was conducted using a relatively limited dataset with constrained environmental variability. Consequently, model performance may decrease under diverse agricultural conditions involving extreme illumination variation, complex backgrounds, multiple disease co-occurrence, and unseen field environments.

Future work will focus on expanding the dataset by increasing the diversity of acquisition conditions, geographical locations, and disease severity levels. Additional groundnut diseases and multiple crop species will also be incorporated to develop a multi-crop disease classification framework.

Declarations

Acknowledgements. The authors gratefully acknowledge Sri Mata Amri- tanandamayi Devi (Amma), Chancellor, Amrita Vishwa Vidyapeetham, for her inspiration and for providing financial support for the Article Processing Charges (APC) of this publication.

Funding. The authors are grateful to the ANRF, New Delhi, India, for funding the project entitled "Light Weight Deep Learning Based Mobile Application for the Early

Detection, Identification, and Spatiotemporal Monitoring of Plant Diseases (Grant No. SUR/2022/004268)."

Data availability

The dataset used in this study is publicly available through the Figshare repository at https://doi.org/10.6084/m9.figshare.33203976.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that no competing financial interests or personal relationships could have influenced the work reported in this paper.

References

1. Tan XL, et al. Bambara Groundnut: An Underutilized Leguminous Crop for Global Food Security and Nutrition. *Front Nutr.* 2020;7:1–16. https://doi.org/10.3389/fnut.2020.601496.

2. Hossain MM, et al. Plant Disease Dynamics in a Changing Climate: Impacts, Molecular Mechanisms, and Climate-Informed Strategies for Sustainable Management. *Springer.* 2024. https://doi.org/10.1007/s44279-024-00144-w.

3. Gupta PK. Current status of Cercosporoid fungi in India, effective management strategies and future directions. *Indian Phytopathol.* 2022;75(2):303–314. https://doi.org/10.1007/s42360-022-00462-3.

4. Buja I, Sabella E, Monteduro AG, Chiriacò MS, De Bellis L, Luvisi A, Maruccio G. Advances in Plant Disease Detection and Monitoring: From Traditional Assays to In-Field Diagnostics. *Sensors.* 2021;21(6):2129. https://doi.org/10.3390/s21062129.

5. Thomidis T, et al. Effects of temperature and leaf wetness duration on pathogens causing preharvest fruit rots on tomato. *J Plant Pathol.* 2023;105(4):1431–1448. https://doi.org/10.1007/s42161-023-01443-9.

6. Hasanaliyeva G, et al. Innovations in Disease Detection and Forecasting: A Digital Roadmap for Sustainable Management of Fruit and Foliar Disease. *Agronomy.* 2022;12(7). https://doi.org/10.3390/agronomy12071707.

7. Barbedo JGA. Plant disease identification from individual lesions and spots using deep-learning. *Biosyst Eng.* 2019;180:96–107. https://doi.org/10.1016/j.biosystemseng.2019.02.002.

8. Shoaib M, et al. An advanced deep learning models-based plant disease detection: A review of recent research. *Front Plant Sci.* 2023;14:1–22. https://doi.org/10.3389/fpls.2023.1158933.

9. Asif M, et al. Technological Progress Toward Peanut Disease Management: A Review. *Sensors.* 2025;25(4):1255. https://doi.org/10.3390/s25041255.

10. Khan N, et al. Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture. *Sustainability.* 2021;13(9):1–31. https://doi.org/10.3390/su13094883.

11. Qadri SAA, et al. Advances and Challenges in Computer Vision for Image-Based Plant Disease Detection. *IEEE Trans Autom Sci Eng.* 2025;22:2639–2670. https://doi.org/10.1109/TASE.2024.3382731.

12. Thanjaivadivel M, Gobinath C, Vellingiri J, Kaliraj S, Femilda Josephin JS. EnConv: enhanced CNN for leaf disease classification. *Journal of Plant Diseases and Protection.* 2025;132(1):32. https://doi.org/10.1007/s41348-024-01033-6.

13. Anbumozhi A, Shanthini A. GLDICCNN Model: Groundnut Leaf Diseases Identification and Classification. *J Adv Inf Technol.* 2024;15(7):812–821. https://doi.org/10.12720/jait.15.7.812-821.

14. Balafas V, et al. Machine Learning and Deep Learning for Plant Disease Classification. *IEEE Access.* 2023;11:114352–114377. https://doi.org/10.1109/ACCESS.2023.3324722.

15. Dheeraj A, Chand S. LWDN: lightweight DenseNet model for plant disease diagnosis. *J Plant Dis Prot.* 2024;131(3):1043–1059. https://doi.org/10.1007/s41348-024-00915-z.

16. Reddy JN, et al. Analysis of Classification Algorithms for Plant Leaf Disease Detection. In: *Proceedings of IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT).* 2019. p. 1–6. https://doi.org/10.1109/ICECCT.2019.8869090.

17. Aishwarya N, et al. Smart farming for detection of tomato plant diseases using lightweight deep neural network. *Multimed Tools Appl.* 2023;82(12):18799–18810. https://doi.org/10.1007/s11042-022-14272-2.

18. Seethapathy P, et al. A Mobile Application for the Early Detection of Dolichos Yellow Mosaic Virus. In: *Proceedings of IEEE ICETA.* 2025. p. 1–6. https://doi.org/10.1109/MPSecICETA64837.2025.11118776.

19. Vallabhajosyula S, Sistla V, Kolli VKK. Transfer learning-based deep ensemble neural network for plant leaf disease detection. *J Plant Dis Prot*. 2022;129(3):545–558. https://doi.org/10.1007/s41348-021-00465-8.

20. Pushpa BR, Yogesh BM, Subhash KB. Corn Plant Disease Detection at Initial Stage Using Deep Learning Models. In: *Proceedings of the 4th International Conference on Sustainable Expert Systems (ICSES)*. 2024. p. 756–763. https://doi.org/10.1109/ICSES63445.2024.10763250.

21. Damaraju GS, et al. PlantHealthAI: An Integrated System for Plant Disease Detection, Severity Prediction With Knowledgebase Chatbot Support. In: *Proceedings of the 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. 2023. p. 1–7. https://doi.org/10.1109/SMARTGENCON60755.2023.10442890.

22. Xu L, et al. Peanut leaf disease identification with deep learning algorithms. *Mol Breed*. 2023;43(4). https://doi.org/10.1007/s11032-023-01370-8.

23. Guo Z, et al. Construction and Validation of Peanut Leaf Spot Disease Prediction Model. *Agronomy*. 2024;14(2). https://doi.org/10.3390/agronomy14020294.

24. Nigar N, et al. Improving Plant Disease Classification Using Explainable AI. *IEEE Access*. 2024;12:100005–100014. https://doi.org/10.1109/ACCESS.2024.3428553.

25. Hassija V, et al. Transformers for vision: A survey on innovative methods for computer vision. *IEEE Access*. 2025. https://doi.org/10.1109/ACCESS.2025.3571735.

26. Zhang T, et al. Cas-ViT: Convolutional additive self-attention vision transformers for efficient mobile applications. *IEEE Trans Image Process*. 2026. https://doi.org/10.1109/TIP.2026.3655121.

27. Rajput SS, et al. Robust apple leaf disease diagnosis for sustainable horticulture: Overcoming background noise with a multilayer transformer-based approach. *Sci Rep*. 2026. https://doi.org/10.1038/s41598-026-45817-2.

28. Remya S, Anjali T, Abhishek S. The Power of Vision Transformers and Acoustic Sensors for Cotton Pest Detection. *IEEE Open J Comput Soc*. 2024;5:356–367. https://doi.org/10.1109/OJCS.2024.3419027.

29. Hu Y, et al. LeafB-Net: Integrating frequency and spatial features with dual-channel attention for robust tomato disease diagnosis. *Pattern Anal Appl*. 2026;29(2):64. https://doi.org/10.1007/s10044-026-01643-w.

30. Ngo D, et al. Edge intelligence: A review of deep neural network inference in resource-limited environments. *Electronics*. 2025;14(12):2495. https://doi.org/10.3390/electronics14122495.

31. Hu L, et al. A paradigm-level evaluation of deep learning architectures for plant pest and disease recognition under data-scarce field conditions. *Front Plant Sci.* 2026;17:1807927. https://doi.org/10.3389/fpls.2026.1807927.

32. Mahmood SA, et al. Lightweight deep learning model-based UAV detection. *Multimed Tools Appl.* 2024;83(32):9881–9902. https://doi.org/10.1007/s11042-024-20328-2.

33. Aishwarya M, Reddy AP. Dataset of groundnut plant leaf images. *Data Brief.* 2023;48:109185. https://doi.org/10.1016/j.dib.2023.109185.

34. Cap QH, et al. LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis. *IEEE Trans Autom Sci Eng.* 2022;19(2):1258–1267. https://doi.org/10.1109/TASE.2020.3041499.

35. Dhingra G, et al. Study of digital image processing techniques for leaf disease detection and classification. *Multimed Tools Appl.* 2018;77(15):19951–20000. https://doi.org/10.1007/s11042-017-5445-8.

36. S. Remya, A. D. Devnandan, S. Gopika, and A. S. Kumar, "Plant disease detection using attention-guided bilinear fusion for edge computing," *Procedia Comput. Sci.*, vol. 283, pp. 267–276, 2026. doi: 10.1016/j.procs.2026.06.114.

37. K. Vinay, V. Surya, S. Thushar, T. Singh, and A. Sahay, "A deep learning framework for early detection and diagnosis of plant diseases," *Procedia Comput. Sci.*, vol. 258, pp. 1435–1445, 2025.