A Lightweight MobileViT-Based Framework for Carambola Leaf and Fruit Disease Detection

S.M. Abdullah Al Muhib, Rejowan Arifin Nayeem, Shalim Shadman Eshan, Jarin Tasnim Showrin, Anisa Khatun Bristy, Shahriar Marjan*, Nafiz Ahmed Emon

Department of Computer Science and Engineering, Daffodil International University, Dhaka-1216, Bangladesh
{muhib15-5107, nayeem15-5108, 252-15-025, showrin15-4547, bristy15-5227, marjan15-5126*, nafiz.cse}@diu.edu.bd

Abstract. Carambola (*Averrhoa carambola*) is one of the most important tropical fruits in terms of economy and nutrition. Production and quality of fruit crop can be severely affected by numerous foliar and fruit diseases. In this research, we propose a complete pipeline establishing a real-world multi-organ dataset of carambola leaves and fruits, evaluate several architectures, including common deep learning models, Vision Transformer-based architectures, and hybrid Transformer-CNN models, train an efficient, mobile-friendly model deployable in field conditions. 2,618 images were acquired in various environments and pre-processed with background cropping, normalization and extensive augmentation for the task of generalization. Multiple model families were trained and tested, and the best performing MobileViT hybrid architecture was able to achieve an overall accuracy of 99.67% along with comparable precision, recall, F1 score performance using less than 0.95 million parameters only. Grad-CAM interpretability analysis further demonstrated that the model successfully highlighted disease-related regions, improving reliability and interpretability. Additionally, the pre-trained model was also implemented into an Android application for on-device disease diagnosis without relying on high-end computers. The major part of this research contributes to precision agriculture by reducing reliance on expert inspection and enabling timely intervention to minimize crop loss while supporting sustainable, data-driven cultivation.

Keywords: Starfruit, Leaf and Fruit Disease, Deep Learning, Vision Transformer, MobileViT, Precision Agriculture.

1 Introduction

Carambola (*Averrhoa carambola*) is a tropical fruit commonly called star fruit and is appreciated for its unique, star-shaped appearance, inviting taste, and nutritious content [1]. It is packed with Vitamin C, antioxidants, and fiber, which provide many health benefits. It is also utilized in traditional medicine as well as having industrial applications. In an age of growing demand, keeping crops healthy and boosting yield are top concerns for farmers and agricultural professionals [2].

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_60

To date, carambola disease monitoring has primarily been conducted by manual inspection by experts, which is intrinsically laborious, time-consuming and error-prone. Visual detection of early-stage infections is challenging, and low-spatial-resolution methods may negatively affect yield losses and fruit quality due to late interventions. Such constraints reinforce the call for automated, reliable and fast disease detection system that helps farmers make correct crop management decisions [3].

Recent developments in computer vision and deep learning have dramatically transformed the field of plant disease diagnosis. Convolutional Neural Networks (CNNs) outperform on leaf-level disease identification, whereas transformer-based architectures such as lightweight Mobile Vision Transformers are suitable for mobile deployment. Hybrid solutions combining CNNs with attention models have also improved the extraction of features and classification performance. However, most of the previous methods are limited in single-organ detection (e.g., only leaf detection), and adopt heavy networks that cannot support real-time field applications. In addition, most models are black boxes which cannot be interpreted and do not generalize well across various environmental conditions and crop organs.

These challenges reveal a pressing gap of the field: missing a unified, lightweight and mobile-friendly framework to detect diseases on both leaves and fruits with high accuracy, interpretability, real-time application. There is additionally a rising need for approaches that can run efficiently on low-resource devices and supply farmers in the fields with relevant feedback.

To alleviate these shortcomings, in this work, we present a large-scale benchmark dataset on carambola leaf and fruit diseases, covering various types of disease categories in practice [4]. Based on this dataset, we further developed a MobileViT based framework for automatic detection and classification using lightweight transformer architecture that is tailored to mobile platforms. The system embeds preprocessing, data augmentation, and explainability tools like Grad-CAM to render the predictions interpretable. Our approach by implementing the trained model in a phone app can provide real-time, on-field disease detection at leaf level from smartphone images. It is expected that this work will promote the development of automated carambola disease detection, and have laid a foundation for multi-organ, mobile-based plant disease diagnosis with features such as good scalability, high efficiency and easy-to-use in smart agriculture.

2 Literature Review

Recent works have investigated deep learning and transformer-based methods for plant disease detection, largely drawing attention to leaf-level classification with CNN and Vision Transformer (ViT) methods. However, their applicability is limited on real-time multi-organ detection for mobile condition.

Datta et al. (2025) [5] proposed a CNN-based model to automatically identify and classify star fruit diseases, achieving an F1-score of 1.0 for carambola detection and 0.98 for disease classification. Although the model had high prediction accuracy, it is time-consuming and less robust across different environments. Possible future work

could explore of lightweight mobile friendly architectures and better generalization across varied environments.

Li et al. (2024) [6] developed PMVT, a small Mobile Vision Transformer technology that combines inverted residual block and CBAM for lightweight plant disease detection on mobile devices. The model also achieved 93.6% on wheat, 85.4% on coffee and rice, respectively, using only 0.98M parameters and outperformed that of MobileNetV3 and SqueezeNet. However, it has been confined to leaf-level detection without cross-organ generalization and disease severity estimation.

Tonmoy et al. (2025) [7] designed MobilePlantViT, a mobile friendly hybrid Vision Transformer which shows 80–99% accuracy over wide-ranged plant disease datasets with with only 0.69M parameters and outperforms MobileViTv1 and v2.

Barman et al. (2024) [8] proposed ViT-SmartAgri, a smartphone embedded Vision Transformer model to diagnose tomato leaf disease with an accuracy of 90.99% over 10,010 images which outperformed Inception V3. Although it had real-time potential, the model's performance tested was limited to tomato leaves and there is no cross-crop generalization.

Parez et al. (2023) [9] introduced GreenViT, a fine-tuned Vision Transformer model designed to enhance the localization of disease regions by directly operating on grids of image patches instead of agglomerations of CNN features. It outperformed CNN-based models on the benchmark datasets in terms of both precision and robustness.

Ding and Yang (2024) [10] have presented a transfer learning model based on MobileViT to predict apple leaf diseases as grey mould, rust, brown spot, scar disease and leaf spot. On a small and complex real-field dataset, the pruned MobileViT achieved 98.54% accuracy, with loss value 0.125 and 2.6 ms prediction time per image (vs. Vision Transformer and Swin Transformer models). The above study has high accuracy, but it is relatively restricted to apple leaves, and the sample size used in the transfer learning category was small.

Khattak et al. (2021) [11] proposed a CNN model to automatically detect diseases (black spot, canker, scab, greening and melanose) in citrus fruit and leaves. On testing with the Citrus and PlantVillage datasets, the model achieved an accuracy of 94.55%, better than other deep learning models in disease classification. But the architecture involves using just CNNs which is computationally intensive and not applicable in real-time or for mobile applications.

Despite the appealing performance of existing works in leaf disease detection, there is still a gap to develop lightweight and flexible models that can work for both leaf and fruit disease analysis which are scalable with field use. The majority of current systems are limited by complex architectures, lack of training data and the inability to implement their functionality in a practical way on mobile devices, thus crippling them for real world application by farmers. To fill this gap, the proposed MobileViT-based system presents a very efficient and human-understandable model which not only achieves high accuracy with good generalization capability but can also be easily embedded within mobile application to recognize disease in real-time. This novel approach offers an inexpensive and user-friendly tool to upgrade accessibility, diagnostic velocity and environmental monitoring in smart agriculture for Averrhoa carambola.

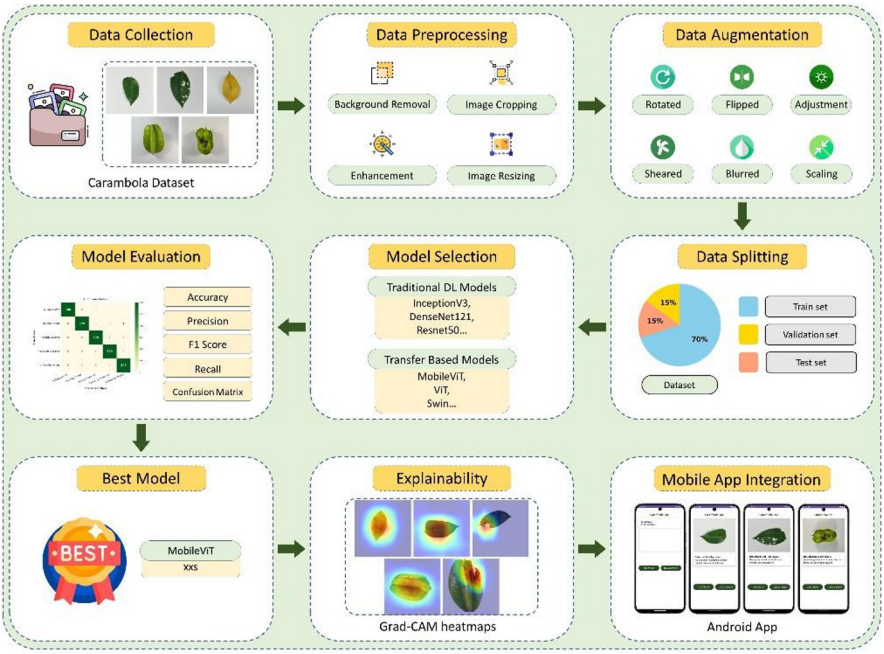

Fig. 1. Overview of the Proposed MobileViT-Based Carambola Disease Detection Pipeline.

3 Methodology

The workflow of the full system for Carambola disease detection proposed was described in Fig. 1. The applied methodology is a pipeline that encompasses stages: dataset preparation, preprocessing, data augmentation, model training, evaluation and deployment. After building a standardized and augmented dataset, several deep learning-transformer models have been trained to receive the best model performance. All models were selected based on comprehensive evaluation with classical performance metrics. The most accurate MobileViT architecture was studied using explainability tools, such as Grad-CAM, and deployed into a mobile application for real-time diagnosis of diseases. This step-by-step approach guarantees that we obtain an efficient, scalable and interpretable pipeline for automated disease detection of Carambola leaves and fruits. Finally, the optimized model was integrated with a mobile app that could be used to detect diseases of Carambola leaves and fruits directly in real-time based on smartphone images, which would offer a practical and easy-to-use tool for farmers and agricultural workers.

3.1 Data Acquisition

A complete dataset of Carambola leaves and fruits was created to assist in automatic disease detection and classification [4]. All images were of high-resolution and were taken periodically between October 2024 and January 2025 in different areas Casing

variability in environment conditions and disease trends. All photos were captured with a natural light source (daylight) using modern smartphones. The dataset contains five separate categories as Insect Hole Leaves, Yellow Leaves, Healthy leaves, Unhealthy Fruits and Healthy Fruits; representing health statuses in both leaves and fruits (Fig. 2). Such a diverse and well-balanced dataset could be used as a good foundation to train and validate deep learning models, which might contribute to the development of precision agriculture and plant disease diagnosis in tropical fruit crops.

Table 1. Statistical Analysis of the Carambola Leaf and Fruit Dataset.

Class Name	Original Images	Augmented Images
Insect Hole Leaves	518	3,000
Yellow Leaves	479	3,000
Healthy Leaves	658	3,000
Unhealthy Fruits	478	3,000
Healthy Fruits	485	3,000
Total	2,618	15,000

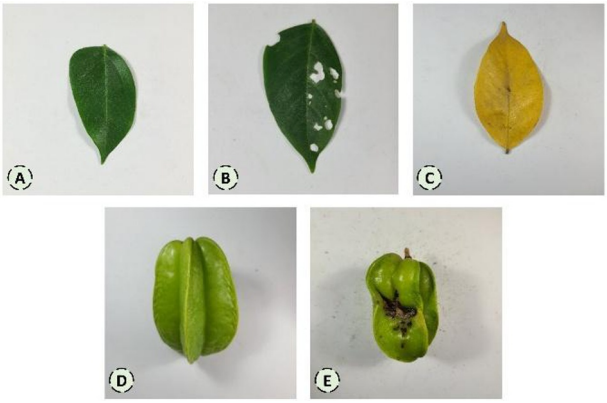

Fig. 2. Sample Images from the Carambola Leaf and Fruit Dataset: (A) Healthy Leaf, (B) Insect Hole Leaf, (C) Yellow Leaf, (D) Healthy Fruit, (E) Unhealthy Fruit

3.2 Data Preprocessing

In order to have uniform data and enhance model performance, a number of image preprocessing operations were performed on the collected Carambola dataset before training. The background of each image was removed to eliminate unnecessary visual noise and emphasize the focus on the target object which is disease-specific features. Brightness and contrast were then adjusted for image normalization to compensate for varying intensity in acquired images. Noise reduction filters were employed in order to avoid visual artifact, enhancing the clarity of images and delineation edges. Lastly, all the images were resized to a common resolution of 224×224 and aspect ratios were

consistent for compatibility with deep learning frameworks as well as uniform input sizes.

3.3 Data Augmentation

The pre-processed Carambola dataset was subjected to extensive data augmentation, in order to increase the model's generalization capacity. We augment to artificially increase the size of the data as well as to maintain class balance across all five classes. Statistical analysis after augmentation is given in Table 1. The transformations applied were: rotation (±30°), horizontal and vertical flipping, brightness and contrast change, shearing, scaling and Gaussian blurring. These operations applied controlled changes in image orientation, lighting and geometry, effectively simulating various light and environmental source variations.

3.4 Model Architecture and Selection

In the arena of agricultural image analysis, choice of an appropriate model architecture is crucial for accurate and efficient disease detection. Conventional deep learning models like VGG, ResNet, Inception and DenseNet have shown powerful potential in hierarchical feature extraction and pattern recognition via plant images. However, such frameworks often require substantial computational resources and enormous training datasets. To overcome this gap, lightweight hybrid architectures such as MobileViT have been proposed which capture local spatial information in convolutions and global contextual information through transformers. A key novelty of employing MobileViT in this study lies in its design for low-resource environments. Compared to conventional transformers, MobileViT significantly reduces memory usage and computational cost through compact patch embeddings and lightweight attention blocks. This allows the model to run efficiently on smartphones with reduced inference latency, making it highly suitable for real-time field conditions where network connectivity and hardware capabilities are limited.

Deep learning models. InceptionV3 is a very deep Convolutional Neural Network based on Inception modules which use multiple convolution filters of various sizes to capture fine as well as coarse visual patterns simultaneously concatenating their outputs. Xception generalizes the Inception concept by replacing standard convolutions with depth-wise separable convolutions altogether to let the model decouple between the spatial and channel-wise feature learning for better performance and efficiency. DenseNet121 connects each layer to every other layer in a feed-forward manner, which fully reuse features and mitigate vanishing gradient by dense connectivity. MobileNet is designed for low computation cost with depth-wise separable convolutions and inverted residuals, which makes it also suitable for mobile/embedded devices, meanwhile achieving competitive accuracy. VGG16 and VGG19 are well-known for having very uniform architecture consisting of only 3×3 convolutions with stacking followed by pooling, with the only difference being the depth (16 or 19 layers) – but also renowned for their clarity and effectiveness despite large number of its parameters. ResNet50

introduces residual functionality which can learn the identity mappings and we believe that this actually helps in very deep model since it avoids vanishing gradient issue and makes convergence faster.

Transfer-based models. Swin Transformer (swin_base_patch4_window7_224) is a hierarchically structured vision transformer, which utilizes shifted windows for efficient image processing and introduces local self-attention for non-overlapping regions while connecting patches across stages to achieve both computation efficiency and scalability towards high-resolution tasks. ViT (vit_base_patch16_224) represents the original Vision Transformer model which considers an image as a sequence of small non-overlapping patches with fixed size, embeds them, and runs it through layers of transformer encoders to capture long-range dependencies over all positions of the input.

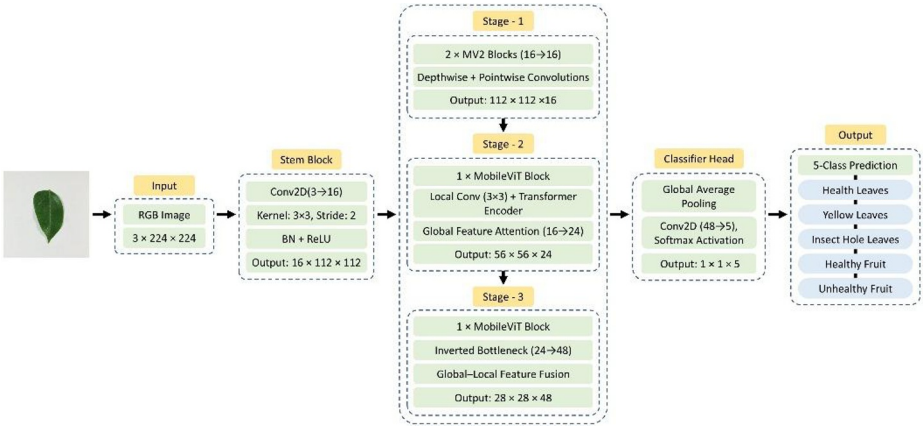

Fig. 3. Architecture of the Proposed MobileViT-Based Carambola Disease Detection Model.

MobileVit. To be specific, MobileViT enhances local pattern extraction by employing CNNs and transformer-style self-attention for representing long-range dependencies. The mobile building block (unit) is a MobileViT block that interleaves local convolutions and transformers for flattened spatial tokens. This block inherits the translation-equivariant inductive bias from convolutions while utilizing attention mechanisms to enhance the global context. The overall network alternates between standard inverted residual and depth-wise separable convolutional stages (for lower-level features and spatial reduction) and multiple MobileViT blocks applied at progressively coarser spatial resolutions. The MobileViT-XXS variant further reduces both width (channel dimensionality) and depth (number of blocks) to meet strict computational and memory constraints while retaining the hybrid CNN-transformer efficiency. The architecture diagram of MobileViT is shown in Fig. 3.

Let $X \in \mathbb{R}^{H \times W \times C}$ be a feature map from previous convolutional layers, where H, W, and C denote height, width, and number of channels, respectively. The MobileViT

block begins with a local convolutional path to generate an enhanced local feature map X_{conv}:

$$X_{conv} = DWConv\big(PWConv(X)\big) \tag{1}$$

After normalization and nonlinearity, the feature map is divided into non-overlapping patches (tokens) of size $p_h \times p_w$. Let $P(\cdot)$represent the rearrangement of the feature map into a sequence of tokens:

$$T_i = P(X_{conv}) \tag{2}$$

$$Ti \in \mathbb{R}^{p_h p_w C} \tag{3}$$

$$N = \frac{HW}{p_h p_w} \tag{4}$$

Each token T_I is linearly projected into a latent dimension d:

$$Z_i = W_p T_i + b_p \tag{5}$$

$$Z_i \in \mathbb{R}^d \tag{6}$$

The transformer encoder processes the sequence $Z_i$through multi-head self-attention (MHSA) and feed-forward networks (FFN). The scaled dot-product attention for one head is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V \tag{7}$$

For MHSA with h heads, the outputs are concatenated and projected:

$$MHSA(Z) = W_0[head_1; \ldots; head_h] \tag{8}$$

A pre-norm residual formulation is applied as follows:

$$Y = Z + MHSA\big(LayerNorm(Z)\big) \tag{9}$$

$$\hat{Y} = Y + FFN\big(LayerNorm(Y)\big) \tag{10}$$

The transformer outputs $\hat{Z}_I$ are reshaped back to spatial form using the inverse patch operator:

$$X_{trans} = P^{-1}\big(\{Z_i\}_{i=1}^{N}\big) \tag{11}$$

$$X_{trans} \in \mathbb{R}^{H \times W \times C'} \tag{12}$$

Local and global features are fused via concatenation and pointwise convolution:

$$X_{fused} = Act(W_f[X_{conv}, X_{trans}] + b_f) \tag{13}$$

$$X_{out} = X + X_{fused} \tag{14}$$

Depth-wise separable convolutions, defined as

$$\mathrm{DSConv(U)} = \mathrm{PWConv}\big(\mathrm{DWConv(U)}\big) \tag{15}$$

significantly reduce computational cost and parameters. During training, standard cross-entropy loss is employed:

$$\mathrm{L_{CE}} = -\sum_{k=1}^{K} y_k \log\left(\frac{e^{sk}}{\sum_j e^{s_j}}\right) \tag{16}$$

The MobileViT-XXS configuration employs inverted residual blocks for early stages, depth-wise separable convolutions for computational efficiency, smaller transformer dimensions with fewer heads, and compact patch sizes to maintain tractable attention complexity $O(N^2 d)$.

MobileViT's design rationale lies in balancing local inductive biases and global dependencies: convolutions efficiently capture local features such as textures and edges, while self-attention models distant spatial relationships. By fusing these representations through pointwise convolution and residual connections, MobileViT achieves robust feature learning with efficient computation. Overall, the MobileViT-XXS variant provides an optimal trade-off between accuracy and resource utilization, combining lightweight convolutional processing with compact global reasoning for efficient visual recognition.

Table 2. Performance Metrics Comparison of the Employed Models

Model	Accuracy	Recall	Precision	F1-Score
MobileViT	99.67%	99.67%	99.67%	99.67%
InceptionV3	96.10%	96.00%	96.00%	96.00%
Xception	95.72%	96.00%	96.00%	96.00%
DenseNet121	94.98%	95.00%	95.00%	95.00%
MobileNet	94.98%	96.00%	96.00%	96.00%
VGG16	90.52%	91.00%	92.00%	91.00%
VGG19	90.33%	93.00%	93.00%	93.00%
ResNet50	71.93%	73.00%	75.00%	72.00%

Table 3. Ablation Study of Transformer-Based Models

Model	Accuracy	Precision	Recall	F1-Score	Parameters
MobileViT	0.9967	0.9967	0.9967	0.9967	952,629
Swin	0.9940	0.9940	0.9940	0.9940	86,748,349
DeiT	0.9940	0.9941	0.9940	0.9940	85,802,501
Mixer	0.9827	0.9828	0.9827	0.9827	59,115,317
ViT	0.9807	0.9810	0.9807	0.9807	85,802,501

3.5 Experimental Setup

The experimental test-bed, for training and evaluation of the Carambola disease detection models was established with a view to maintain reproducibility and efficient usage of computing power. The experiments were implemented using Python in the programming environment VS Code.

We loaded the dataset into memory with a PyTorch Dataset Class in PyTorch and resized all images to 224×224 pixels. The batch size of model training is set to 16 and models are trained with learning rate of 1×10^{-4} for 20 epochs. The parameter updates were done by Adam optimizer and the cross-entropy loss was used as objective function. The training was performed using GPU acceleration when available, for multi-GPU training we used DataParallel.

Standard 70:15:15 train/validation/test splits were generated with stratified sampling to ensure that the two pathological classes would be balanced across subsets. We monitored model performance on the training loss and accuracy while training and used evaluation on the validation set to select the models. After training, the models were tested with test set based on standard metrics of accuracy, precision, recall, F1-score and confusion matrix and ROC curves. Finally, Grad-CAM heatmaps were used to understand the model's predictions.

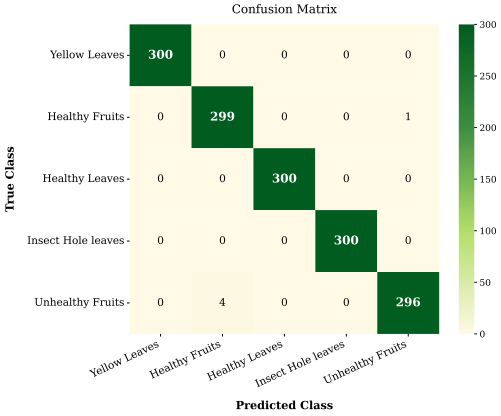

Fig. 4. Confusion Matrix of the Proposed MobileViT Model.

4 Result and Discussion

4.1 Optimal Model Performance

The MobileViT (mobilevit_xxs. cvnets_in1k) model achieved excellent performance in detecting and classifying carambola leaf or fruit diseases. All in all, the model was able to make very reliable and consistent predictions for all five classes with an overall accuracy of 99.67%. The confusion matrix (Fig. 4), which suggests low misclassification of the samples and significance of the model. Furthermore, the ROC curves (Fig.

5) of all categories were 1.00, indicating a very good classification performance and no class imbalance problem in the prediction procedure. The loss-accuracy plots for training and validation (Fig. 6) exhibit consistent convergence without overfitting and underfitting, indicating the powerful generalization of the model. Overall, the MobileViT model was determined to be a most suitable architecture for this study because of its excellent trade-off between accuracy, efficiency and generalization in application of carambola leaf and fruit disease detection system.

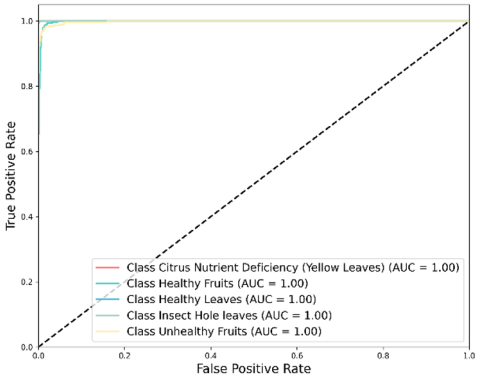

Fig. 5. ROC Curves for the MobileViT Model Across All Classes.

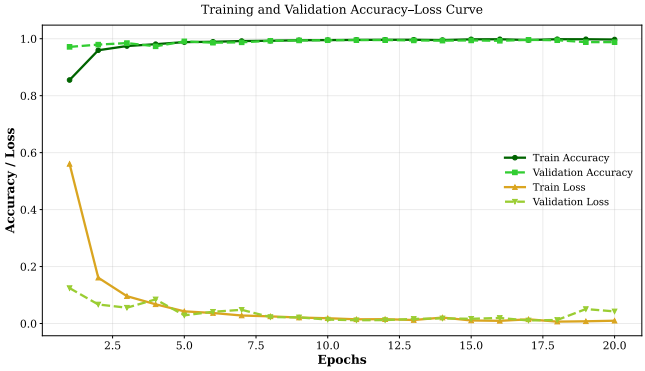

Fig. 6. Training and Validation Loss–Accuracy Curves of the MobileViT Model.

4.2 Comparative Evaluation

Comparing to other deep learning models as summarized in Table 2 the MobileViT XXS architecture is outperforming the CNN-based architectures. As illustrated in Fig. 7, MobileViT obtained the highest validation accuracy, precision, recall and F1-score among all models tested indicating its stability and generality on carambola leaf as well as fruit disease detecting. In comparison with typical models e.g., VGG, ResNet, and DenseNet that heavily depend on convolutional feature propagation, MobileViT

effectively combines local spatial learning from CNNs and global context-sensitive learning from NSF. Such hybrid design is helpful for the model to capture finer disease patterns more accurately and generalize better. In addition to, among other models, MobileViT was also the best performing and computationally efficient model which is highly recommended for real-time agricultural applications and mobile deployment in particular.

4.3 Ablation and Comparative Study

The ablation and comparative studies are outlined in Table 3, shown in Fig. 8, compares the performance of the newly pro- posed MobileViT XXS model with state-of-the-art vision transformer-based architectures. Although other models (e.g., Swin Transformer, DeiT, Mixer, and ViT) have similar accuracy and F1-scores to MobileViT XXS, their parameter counts are multiple orders of magnitude large than those in MobileViT XXS. Even with only ~953K parameters, MobileViT XXS was able to produce slightly better result, indicating the gains from bigger transformer models are not as significant in this particular task. This means that the local and global learning of MobileViT hybrid scheme can efficiently learn discriminative features from carambola leaf and fruit images. The extreme low computational cost and fast deployment nature of MobileViT XXS becomes particularly desirable for implementation in the resource-limited settings such as mobile apps for real-time disease prediction without comprising performance or generalization ability.

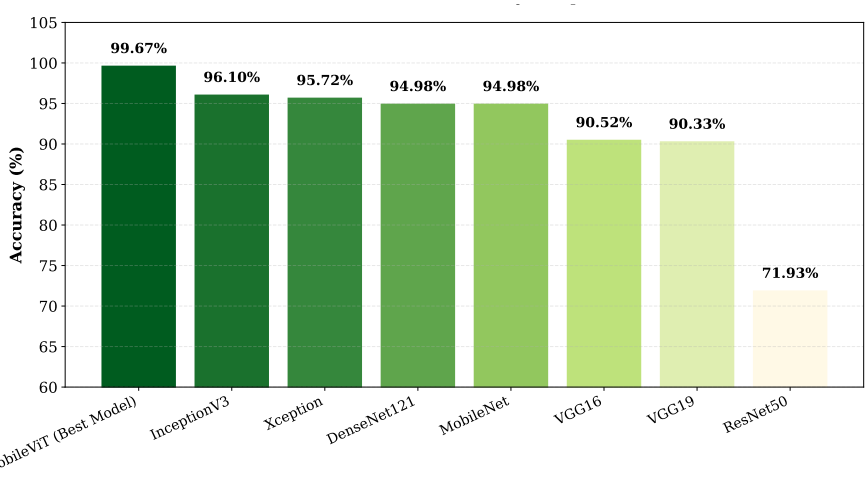

Fig. 7. Validation Accuracy Comparison Across Different Models.

4.4 Explainable Analysis

The explainable investigation into the proposed MobileViT model was conducted with Grad-CAM, which is a visualization technique to gain insight into what features of input data does the network use for making its prediction. As shown in Fig. 9, the Grad-

CAM results correctly pinpoint disease-damaged areas in carambola leaves and fruits, indicating that the model values important features such as injury, blight and insect pest. These visualizations serve to confirm the model's predictions and also aid in interpretability, enabling users of the automated system for disease detection to better comprehend and trust the model. The performance is very promising, demonstrating that MobileViT can well capture the critical regions of each disease class and remain a general model interpretability.

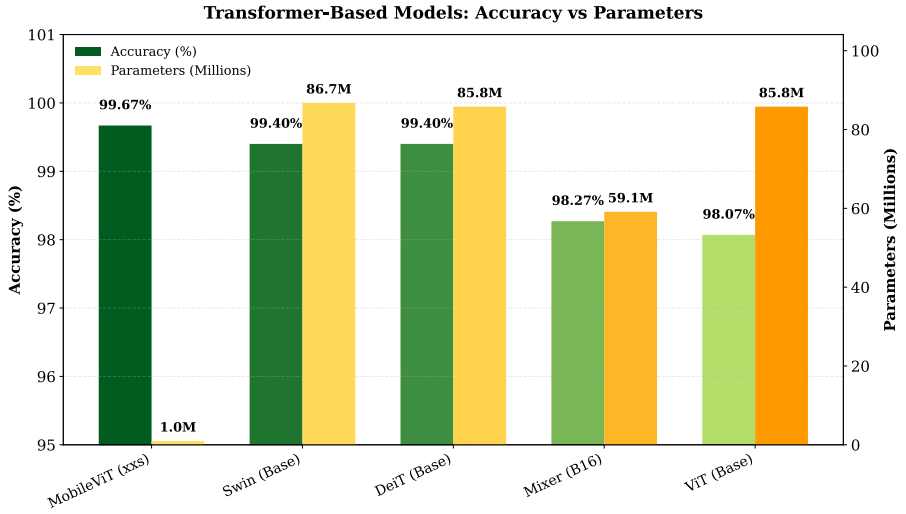

Fig. 8. Accuracy vs. Parameters Comparison of Transformer-Based Models.

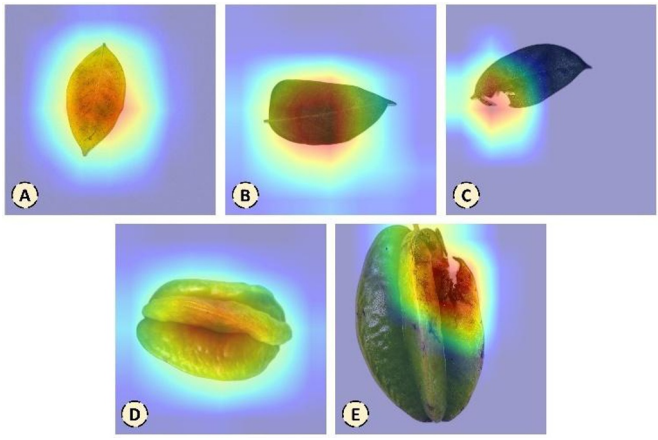

Fig. 9. Grad-CAM Visualization of Disease-Related Regions: (A) Yellow Leaf, (B) Healthy Leaf, (C) Insect Hole Leaf, (D) Healthy Fruit, (E) Unhealthy Fruit

5 Deployment

The trained MobileViT XXS model has been successfully deployed in a mobile application developed using Kotlin and Android Studio, enabling real-time carambola leaf and fruit disease detection. The application allows users to either capture images directly through the device camera or select images from the gallery. Once an image is provided, the integrated model processes it and delivers an immediate classification output, indicating the specific disease or health status of the leaf or fruit. The sample interface of the application is illustrated in Fig. 10, showing the user-friendly layout and interaction flow. This deployment demonstrates the practical utility of the proposed system, providing an accessible, lightweight, and accurate tool for farmers and agricultural practitioners to monitor and diagnose carambola diseases directly in the field.

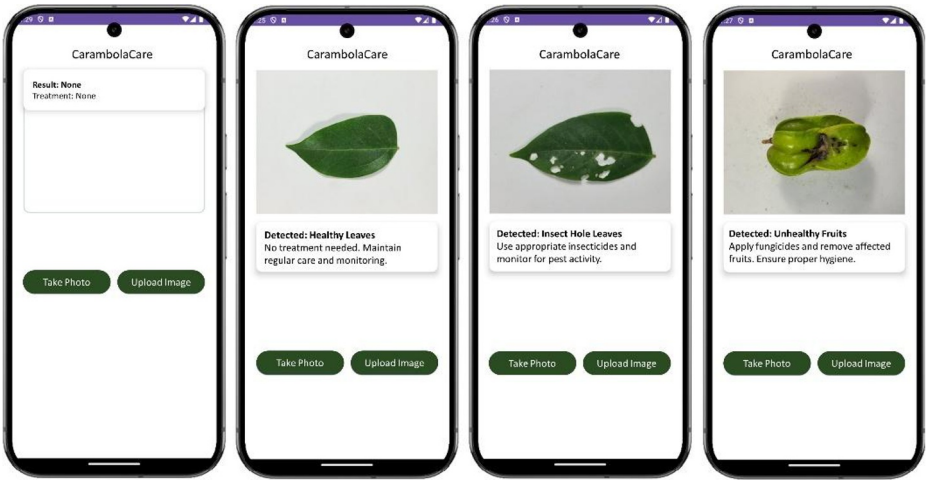

Fig. 10. Mobile Application Interface for On-Device Disease Diagnosis

6 Conclusion

In summary, this work offers a promising basis for intelligent and automatic disease diagnosis in carambola planting, as well as the practical value of deep learning-based techniques on agricultural image analysis. The experimental results have shown that compact and high-performance models can be properly formulated ready for real-world use, reconciling the gap between research accuracy and field practicality. The incorporation of explainability methods also guarantees that the proposed framework is reliable and interpretable for agricultural practitioners, in addition to its high accuracy. In addition to the application of this method itself, we exhibit evidence that low-cost vision-based systems can transfer crop monitoring from a manual task described by human scales into a scalable one, driven by data. As an extension of this work, inclusion of the multi-spectral images, across crop datasets and federated learning can be used to

improve the robustness, adaptability and scalability of models for a wider range of agricultural applications.

References

1. Lakmal, K., Yasawardene, P., Jayarajah, U., Seneviratne, S.L.: Nutritional and medicinal properties of Star fruit (Averrhoa carambola): A review. Food Science & Nutrition 9(3), 1810–1823 (2021). https://doi.org/10.1002/fsn3.2135
2. Harsh, D.: Challenges of Star Fruit Averrhoa carambola: A Comprehensive Overview. International Journal of Science and Research (IJSR) 12(8), 2132–2135 (2023). https://doi.org/10.21275/sr23823113638
3. Mohd Suhaimi, N.I., Mat Ropi, A.A., Shaharuddin, S.: Safety and quality preservation of starfruit (Averrhoa carambola) at ambient shelf life using synergistic pectin–maltodextrin–sodium chloride edible coating. Heliyon 7(2), e06279 (2021). https://doi.org/10.1016/j.heliyon.2021.e06279
4. Muhib, A., Nayeem, R.A., Mezi, N., Emon, N.A.: A Comprehensive Image Dataset for Carambola Leaf and Fruit Disease Classification and Quality Assessment. Data in Brief 60, 111679 (2025). https://doi.org/10.1016/j.dib.2025.111679
5. Datta, S., Banerjee, S., Jana, S.: Disease Classification of Star Fruit (Averrhoa carambola L.) Using Deep Learning. In: Lecture Notes in Networks and Systems, pp. 13–22 (2025). https://doi.org/10.1007/978-981-96-5822-0_2
6. Li, G., Wang, Y., Zhao, Q., Yuan, P., Chang, B.: PMVT: A lightweight vision transformer for plant disease identification on mobile devices. Frontiers in Plant Science 14 (2023). https://doi.org/10.3389/fpls.2023.1256773
7. Rahman, T.M., Hossain, M.M., Dey, N., M.F., M.M.: MobilePlantViT: A Mobile-friendly Hybrid ViT for Generalized Plant Disease Image Classification. arXiv preprint (2025). https://arxiv.org/abs/2503.16628
8. Barman, U., et al.: ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture. Agronomy 14(2), 327 (2024). https://doi.org/10.3390/agronomy14020327
9. Parez, S., Dilshad, N., Alghamdi, N.S., Alanazi, T.M., Lee, J.W.: Visual Intelligence in Precision Agriculture: Exploring Plant Disease Detection via Efficient Vision Transformers. Sensors 23(15), 6949 (2023). https://doi.org/10.3390/s23156949
10. Ding, Y., Yang, W.: Classification of apple leaf diseases based on MobileViT transfer learning. In: International Conference on Image Processing and Artificial Intelligence (ICIPAI 2024), p. 64 (2024). https://doi.org/10.1117/12.3035225
11. Khattak, A., et al.: Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model. IEEE Access 9, 112942–112954 (2021). https://doi.org/10.1109/access.2021.3096895