Vol. 06, No.02x, pp. 219 –230 (2025)
ISSN: 2708-0757

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

AppleVit: A Smart Agricultural Software for Apple Leaf Disease Detection Using AI

Pradeep Gupta*, Rakesh Singh Jadon

Department of CSE, Madhav Institute of Technology & Science, Gwalior (M.P.), India. gupta.pradeep85@gmail.com, rsjadon@mitsgwalior.in
** Correspondance: gupta.pradeep85@gmail.com*

Abstract

Apple leaf diseases endanger global apple production at such an intensity that it demands precise detection systems to control disease spread effectively. Traditional inspection methods and Convolutional Neural Network (CNN)-based models face challenges when processing extended image dependencies in leaf images, which subsequently affects their ability to identify diseases accurately. This research develops AppleViT, a lightweight Vision Transformer (ViT)-based model that applies Vision Transformer technology with self-attention approaches to enhance leaf disease classification accuracy and feature extraction within apple leaf detection systems. AppleViT was trained using a public dataset comprising 9,714 apple leaf images, categorized into four classes: Apple Scab, Black Rot, Cedar Apple Rust, and Healthy. The accuracy rate of AppleViT reached 97.8%, which exceeded the ResNet-50 and EfficientNet-B3 and MobileNetV3 models while operating with 1.3 million parameters suitable for precision agriculture real-time usage. The proposed approach demonstrates both high generalization skills alongside precise precision and recall value measurements for disease categories. Future research will create attention visualization features and mobile application compatibility before expanding the architecture to identify multiple diseases across different plant types. AppleViT highlights the potential of Vision Transformer (ViT) technology as a powerful tool to revolutionize plant disease detection for improving crop yield and disease management worldwide.

Keywords: *Apple Leaf Disease Classification, Deep Learning, Vision Transformer, Self-Attention Mechanism, Precision Agriculture, Computer Vision, Machine Learning.*

Received: July 05th, 2025 / Revised: August 20th, 2025 / Accepted: September 02nd, 2025 / Online: September 06th, 2025

I. INTRODUCTION

The global need for food security depends heavily on apple cultivation due to its important delivery of essential nutrients including vitamins as well as minerals and antioxidants [1]. Sustainable apple cultivation struggles from leaf diseases that result in major harvest losses unless proper disease control methods exist[2].The three pathogenic microbes Botryosphaeria obtusa and Venturia inaequalis together with Gymnosporangium juniperi-virginianae are known to cause severe damage through blemishes and lesions which disrupt photosynthetic functions and fruit development [3]. Growers need correct disease identifications at right times to activate proper control methods and limit economic damage [4].

A. Limitations of Traditional Disease Identification

Crop disease identification has traditionally relied on physical inspection by agricultural experts. While feasible for small-scale farming, these methods demand excessive human labor and time, and they are prone to errors [5]. Early symptom detection is challenging because visible symptoms often appear only after disease progression. Moreover, disease symptom expression can vary with environmental conditions, making consistent classification difficult [6].

B. Apple Leaf Disease Classification using Deep Learning

Automated deep learning methods, particularly Convolutional Neural Networks (CNNs), have revolutionized plant disease identification [7]. Apple leaf disease detection becomes successful with the application of three widely used CNN-based models including ResNet [8], EfficientNet [9] and MobileNet [10]. These models automatically learn relevant image features and outperform traditional machine learning classifiers. However, CNNs have fundamental constraints in capturing long-range dependencies in images. The localized receptive fields of CNNs make it difficult to learn global relationships between distant leaf regions [11]. Pooling operations may further reduce fine-grained spatial details,

doi: 10.38094/jastt62315

lowering accuracy for complex or overlapping disease patterns. Moreover, CNNs typically require large labeled datasets and high computational resources, limiting their deployment in real-time, resource-constrained agricultural environments.

C. Vision Transformers (ViTs) for Image Analysis

Vision Transformers (ViTs) have emerged as a promising alternative, employing self-attention mechanisms to capture both local and long-range dependencies in images. ViTs divide an image into patches and process them using global self-attention, thereby retaining holistic contextual information—something CNNs struggle to achieve [12]. Preliminary research suggests ViTs can improve apple leaf disease detection, but most existing ViT models remain computationally heavy and are rarely optimized for deployment in agricultural edge environments. Recent works have proposed lightweight and hybrid ViT models that integrate CNN-based feature extraction to improve efficiency [13] however, these typically exceed 5–10M parameters, which still poses a challenge for mobile deployment.

D. Proposed AppleViT Model

To address these limitations, this study proposes AppleViT, a lightweight Vision Transformer model for apple leaf disease classification that achieves 97.8% accuracy with only ~1.3M parameters. AppleViT is designed to:

1) *Overcome CNN limitations by leveraging global feature learning through self-attention,*
2) *Enhance classification accuracy by capturing long-range dependencies while maintaining low computational overhead,*
3) *Enable real-time deployment on resource-constrained devices, and*
4) *Provide interpretability via Layer-wise Relevance Propagation (LRP) and attention maps for transparent decision-making.*

Unlike prior works, AppleViT is explicitly optimized for both high performance and mobile feasibility, making it suitable for in-field agricultural applications. By integrating efficient transformer-based techniques with interpretability, AppleViT offers a scalable, transparent, and computationally efficient solution for apple leaf disease detection. This contributes to sustainable apple farming by enabling earlier and more accurate interventions, ultimately supporting global food security.

II. Related Work

Global apple production relies on effective classification and management strategies for apple leaf diseases. Early research in this domain started with manual inspection and rudimentary image processing, but it has now shifted to advanced deep learning methods. Below, we review the progression from traditional techniques to modern deep learning and transformer-based approaches.

A. Traditional Approaches

Early plant disease classification methods used hand-crafted features such as color segmentation, texture analysis, and edge detection[14]. These approaches had limited success, as they did not generalize well to diverse conditions and required expert knowledge to extract features. Classical machine learning classifiers (Support Vector Machines, Random Forests, k-Nearest Neighbors) were applied with these features, but they demanded extensive manual feature engineering and proved difficult to scale beyond small datasets.

B. CNN-Based Approaches

The advent of convolutional neural networks (CNNs) enabled automatic feature extraction for plant disease detection, yielding higher accuracy and rendering manual feature engineering obsolete. Numerous studies have confirmed the effectiveness of CNNs: for example, an ensemble of CNN and Vision Transformer (ViT) achieved 96% accuracy in olive leaf disease detection [15]; a CNN with GAN-based augmentation improved tomato leaf disease classification [16]; and a compact CNN (RegNet) outperformed other models in apple leaf disease detection [17]. A data augmentation approach using background removal for apple leaf disease classification with MobileNetV2 was proposed, highlighting the importance of background preprocessing in improving model robustness under real-world conditions [13].Despite their success, CNNs have limitations in capturing global dependencies due to local receptive fields and in retaining fine-grained spatial details due to pooling layers.

C. ViT-Based Approaches

Recent studies have explored Vision Transformers (ViTs) for plant disease classification, leveraging self-attention to improve feature representation by retaining both global and local context. Notable examples include the use of a MaxViT transformer model achieving ~97% accuracy on tomato leaf diseases [18], an attention-based ViT mapping approach that outperformed CNNs with test accuracies of 85.9%, 89.2%, and 94.2% on different plant disease datasets [19], and an SEViT model which improved fine-grained plant disease classification accuracy [20]. These works demonstrate the strength of ViTs in modeling long-range interactions. However, vanilla ViTs can be computationally heavy, making them less suitable for real-time agricultural applications without modifications or efficiency improvements.

D. Hybrid CNN -ViT Based Approaches

Given the complementary strengths of CNNs and ViTs, researchers have developed hybrid architectures that combine both. For instance, SLViT, a shuffle-convolution-based lightweight ViT, was introduced by integrating a CNN stem with transformer blocks. SLViT demonstrated improved speed, reduced model size, and high precision on benchmark leaf disease datasets such as PlantVillage and a sugarcane leaf disease dataset, highlighting the benefit of hybrid designs [21]. PlantXViT, a CNN–ViT hybrid model, achieved high accuracies (93–98%) on apple, maize, and rice leaf disease datasets [22]. Former Leaf, an efficient ViT model optimized for cassava leaf disease detection through attention pruning and sparse operations, further improved efficiency for crop-specific tasks [23]. AppViT, a hybrid model stacking CNN convolutional blocks with ViT blocks, achieved 96.38% precision on the challenging Plant Pathology 2021 apple leaf

dataset with only ~1.3 million parameters, outperforming ResNet-50 and EfficientNet-B3 by 11.3% and 4.3% respectively, underscoring the power of lightweight hybrid transformers [24]. Similarly, MobilePlantViT, a mobile-friendly hybrid ViT model with only 0.69M parameters, achieved 80–99% accuracy across diverse crop disease datasets [25]. TrIncNet, a transformed Inception-ViT network, was also introduced as a lightweight ViT-based model for crop disease identification [26]. In parallel, TinyResViT, a hybrid of ResNet and ViT designed for on-device corn leaf disease detection, demonstrated efficient performance in real-world field settings [27]. From a data augmentation perspective, InViT-Mixup, a convolutional ViT with Mixup augmentation, showed improved classification accuracy on tomato leaf diseases [28].

Additionally, recent works from 2025 further demonstrate the importance of lightweight and hybrid ViTs in agriculture:

- ViT integration with spectral imaging was demonstrated for precision crop monitoring [29]
- A ViT-based explainable AI framework for maize leaf disease classification achieved 94.97% accuracy with only 1.22 million parameters, enabling real-time mobile deployment and enhancing model transparency through explainable AI [30].
- A hybrid deep learning model for maize leaf disease classification with explainable AI, combining convolutional feature extraction with transformer-based attention layers. Their approach achieved high accuracy while offering interpretability through Grad-CAM visualizations, demonstrating the potential of hybrid architectures in balancing efficiency, accuracy, and transparency in agricultural AI [31].

These examples illustrate a clear trend toward lightweight and hybrid ViT models that balance CNN's inductive biases with Transformers' global attention, achieving high accuracy with fewer parameters.

Our proposed AppleViT aligns with this trend but is tailored specifically for apple leaves, introducing an attention-based architecture with only 1.3M parameters and adding interpretability features (Layer-wise Relevance Propagation) not seen in prior models. This focus on apple-specific disease patterns and model explainability distinguishes AppleViT from similar-accuracy models like SLViT and PlantXViT. A comparative summary of related studies is presented in Table I.

E. Research Gaps

Despite the progress in plant disease AI models, multiple gaps remain. Many studies prioritize highest accuracy on closed datasets, sometimes at the expense of practical considerations like model interpretability and robustness. The opaque "black-box" nature of ViTs and deep models makes it hard for farmers to trust their outputs. Without clear explanations of how decisions are made, user acceptance in agricultural practice is limited. Additionally, most models are trained on limited datasets (often with lab-controlled images), so they risk overfitting and may not generalize to new diseases or field conditions (e.g., different lighting, backgrounds). Thus, a need exists for models that maintain high accuracy and provide transparent reasoning and reliable performance in diverse real-world scenarios. To address these challenges, we propose AppleViT, a lightweight Vision Transformer tailored for apple leaf disease classification. AppleViT leverages self-attention for long-range feature modeling, integrates interpretability mechanisms, and achieves state-of-the-art accuracy with minimal parameters.

F. Contributions of AppleViT

The key contributions of this work are:

1) Lightweight Vision Transformer (AppleViT): A novel ViT-based model with only 1.3M parameters, achieving 97.8% accuracy in apple leaf disease classification.
2) Interpretability: Integration of Layer-wise Relevance Propagation (LRP) and attention visualizations to highlight decision-relevant regions and improve model trust.
3) Generalization: Robust performance across real-world environments through augmentation, transfer learning, and external validation on the Plant Pathology 2021 dataset.
4) Practical Deployment: Demonstration of AppleViT's competitiveness with state-of-the-art CNNs at lower computational cost, underscoring its suitability for real-time mobile and edge applications.

By addressing these gaps, AppleViT provides an interpretable, efficient, and scalable framework for precision agriculture and plant disease management.

III. Materials And Methodology

A. Dataset Description

We utilized a dataset of 9,714 apple leaf images to develop and evaluate the AppleViT model. The images are categorized into four classes: Apple Scab (2,016 images), Black Rot (1,987 images), Cedar Apple Rust (1,760 images), and Healthy leaves (2,008 images). These high-resolution images were collected from varied environments (both lab and orchard settings) to ensure diversity in background and lighting. Prior to training, all images underwent preprocessing to improve generalization. The preprocessing steps included:(a) Data Augmentation – random rotations, flips, scaling, and brightness adjustments were applied to simulate different angles and lighting conditions; (b) Normalization – pixel values were standardized (zero mean, unit variance) per channel; and (c) Dataset Splitting – the dataset was divided into 80% for training and 20% for validation/testing. Table II provides the detailed class-wise split. Sample images from each category, illustrating the kind of leaf and background variations, are shown in Figure 1.

TABLE I. PLANT LEAF DISEASE DETECTION COMPARISON OF PREVIOUS WORK

Reference	Year	Dataset	Methodology	Accuracy	Limitation
[13]	2024	Apple leaves	CNN (lightweight)	98.72% (Precision)	Focused only on apples; limited generalization
[15]	2022	Olive leaves (olive leaf disease)	CNN + ViT ensemble	96%	Limited to olive leaf disease
[16]	2022	Mixed plant leaves	Deep CNN (with GAN augmentation)	High (improved)	Requires data augmentation for best performance
[20]	2022	(Not specified)	SEViT (Squeeze-and-Excitation ViT)	88.34%	Requires large-scale computing
[22]	2022	Apple, maize, rice leaves	PlantXViT (CNN–ViT hybrid)	93.55–98.33%	Limited dataset diversity
[23]	2023	Cassava leaves (not specified)	FormerLeaf (efficient ViT)	(Not reported)	Optimization strategy unclears
[29]	2025	Multi-spectral crop images	ViT with spectral imaging integration	~95%	Requires specialized spectral hardware
[32]	2022	Various plant diseases & pests	ViT-based automated pest identification	96.71%	High computational demand
[33]	2023	Mango leaves	Federated learning-based CNN	97–98%	Requires extensive data for federation
[34]	2023	Citrus, cucumber, grape, tomato leaves	DLMC-Net (deeper lightweight CNN)	92.34–99.50%	Performance varies with lighting conditions
[35]	2025	PlantVillage subset (multi-crop)	Mobile-friendly hybrid ViT	~95%	Edge deployment tested, but dataset limited
[36]	2025	Corn leaves	Lightweight ResNet + ViT hybrid	~94%	Evaluation limited to a single crop
[37]	2024	Multiple plant species	CNN–ViT hybrid with Mixup augmentation	93–97%	Needs large-scale augmented data

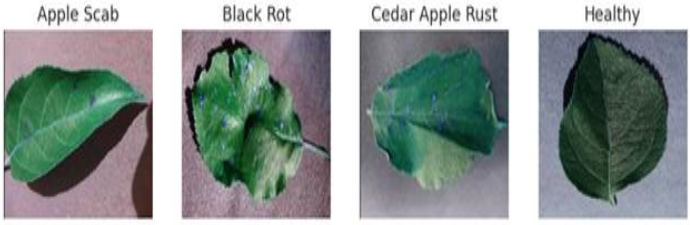

Fig. 1. Sample images of apple leaves from each disease category in the dataset.

TABLE II. PARTITION OF THE APPLE LEAF DATASET INTO TRAINING, VALIDATION, AND TESTING SETS..

Class	Training	Validation	Testing
Apple Scab	2016	453	51
Black Rot	1987	447	50
Cedar Apple Rust	1760	396	47
Healthy	2008	451	51

This balanced partition ensures that the model learns distinct patterns associated with each leaf disease class, reducing bias toward any single category. The presence of diverse backgrounds (e.g., natural foliage vs. plain backgrounds) in the training set helps AppleViT remain robust to real-world scenarios.

B. Overview of the Proposed Model

The proposed **AppleViT** model is a Vision Transformer (ViT)-based architecture designed for apple leaf disease classification. It comprises three major components:

1) Patch Embedding Module

2) Stacked Transformer Encoder Blocks

3) Classification Head

The following subsections detail each component and present the corresponding mathematical formulations.

1) Patch Embedding: Given an input image $X \in \mathbb{R}^{H\times W\times 3}$ of height H, width W, and three RGB channels, the image is divided into non-overlapping square patches of size $P \times P$. The total number of patches is calculated using *(*Eq. 1*)*:

$$N = \frac{H \times W}{P^2} \qquad (1)$$

Each patch x_i $(i = 1,2, \dots, N)$ is flattened into a vector and projected to a D-dimensional embedding space using a trainable matrix $E \in \mathbb{R}^{P^2 \times D}$ as in (Eq.2):

$$z_i^0 = E \cdot \text{flatten}(x_i) + E_{\text{pos},i} \quad (2)$$

where $E_{\text{pos},i}$ is the positional embedding of the i-th patch.

A learnable class token $z_{\text{cls}}^0 \in \mathbb{R}^D$ is appended to the sequence of patch embeddings, and positional embeddings E_{pos} are added as in (Eq.3):

$$Z^0 = [\, z_{\text{cls}}^0;\ z_1^0;\ z_2^0;\ \dots;\ z_N^0\,] + E_{\text{pos}} \quad (3)$$

Here, $Z^0 \in \mathbb{R}^{(N+1)\times D}$ is the initial token sequence for the transformer encoder.

2) Transformer Encoder Blocks: The sequence Z^0 is processed by L identical transformer encoder layers. Each layer consists of:

a) Multi-Head Self-Attention (MHSA): In the ℓ-th layer, the input $Z^{\ell-1}$ is linearly projected into queries Q, keys K, and values V for each attention head as in (Eq.4):

$$Q = Z^{\ell-1}W_Q, K = Z^{\ell-1}W_K, \quad V = Z^{\ell-1}W_V \quad (4)$$

where $W_Q, W_K, W_V \in \mathbb{R}^{D\times d_h}$ are learnable projection matrices, and d_h is the head dimension ($D = h \times d_h$, with h = number of heads).

Attention weights are computed as in (Eq.5):

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_h}}\right) \quad (5)$$

where $A \in \mathbb{R}^{(N+1)\times(N+1)}$ contains pairwise attention scores. The head output is obtained as in (Eq. 6):

$$\text{head}_h = A \cdot V \quad (6)$$

Outputs from all heads are concatenated and projected through $W_O \in \mathbb{R}^{D\times D}$.

b) Add & Norm: The MHSA output is added to the input sequence via a residual connection, followed by Layer Normalization ((Eq. 7):

$$Z' = \text{LN}\big(Z^{\ell-1} + \text{MHSA}(Z^{\ell-1})\big) \quad (7)$$

c) Feed-Forward Network (FFN): The FFN transformation is defined in (Eq. 8):

$$\text{FFN}(x) = W_2\,\sigma(W_1 x + b_1) + b_2 \quad (8)$$

where $W_1, W_2 \in \mathbb{R}^{D\times D}$, b_1, b_2 are biases, and σ is the GELU activation.

The FFN output is added to Z' and normalized (Eq. 9):

$$Z^\ell = \text{LN}(Z' + \text{FFN}(Z')) \quad (9)$$

This output becomes the input for the next transformer layer.

After L layers, the final representation is:

$$Z^L = [\, z_{\text{cls}}^L; z_1^L; z_2^L; \dots; z_N^L\,]$$

where z_{cls}^L contains the global image representation.

3) Classification Head: The class token output is passed through a fully-connected layer to obtain prediction logits (Eq. 10):

$$\hat{y} = \text{softmax}(W_C^T z_{\text{cls}}^L + b_C) \quad (10)$$

where $W_C \in \mathbb{R}^{D\times C}$, $b_C \in \mathbb{R}^C$, and C is the number of classes ($C = 4$ in this study).

The predicted class is the index j with the highest probability:

$$\text{class} = \underset{j}{\text{argmax}}\,\hat{y}_j$$

meaning the class with the largest $\hat{y}_j$ is selected as the model's prediction.

The AppleViT architecture effectively captures long-range dependencies between image regions using self-attention, while maintaining a compact parameter count (~1.3M). Figure 2 illustrates the complete pipeline, including:

- Patch Embedding
- Stacked Transformer Encoders with MHSA + FFN
- Classification Head

C. Training Algorithm

We outline the training and inference procedure for AppleViT in **Algorithm 1**. This includes preprocessing, model training, validation, and deployment steps for clarity.

Algorithm 1: Apple Leaf Disease Detection using AppleViT
Input: Labeled Apple leaf image I, number of disease classes N.
Output: Predicted disease class label $\hat{y}$.

1. Preprocessing: Resize I to 224×224, apply data augmentation (random rotations, flips), and normalize pixel values:
$$I_{norm} = \frac{I' - \mu}{\sigma}$$
2. Model Initialization: Load pre-trained ViT model, freeze layers (if applicable), and replace the final classification head with:
$$q\ \hat{y} = softmax(W_c Z_{cls} + b_c)$$
3. Training: Train the model for E epochs with batch size B. The forward pass extracts ViT features:
$$Z = f_{ViT}(I_{norm}; \Theta_{ViT})$$
Optimize parameters using categorical cross-entropy loss: $L = -\sum_{k=1}^{N} y_k \log \hat{y}_k$ and update using Adam optimizer: $\Theta_{\text{ViT}}^{t+1} = \Theta_{\text{ViT}}^{t} - \eta \frac{\partial L}{\partial \Theta}$
4. Validation: Evaluate performance using accuracy: $\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$ Apply early stopping or learning rate scheduling if necessary.
5. Inference: For a new test image I_{test}, compute class probabilities: $P(y_k \mid I_{\text{test}}) = f_{\text{ViT}}(I_{\text{test}}; \Theta_{\text{ViT}})$ The predicted class is assigned as: $\hat{y} = \underset{k}{\text{argmax}} P(y_k \mid I_{\text{test}})$

End Algorithm

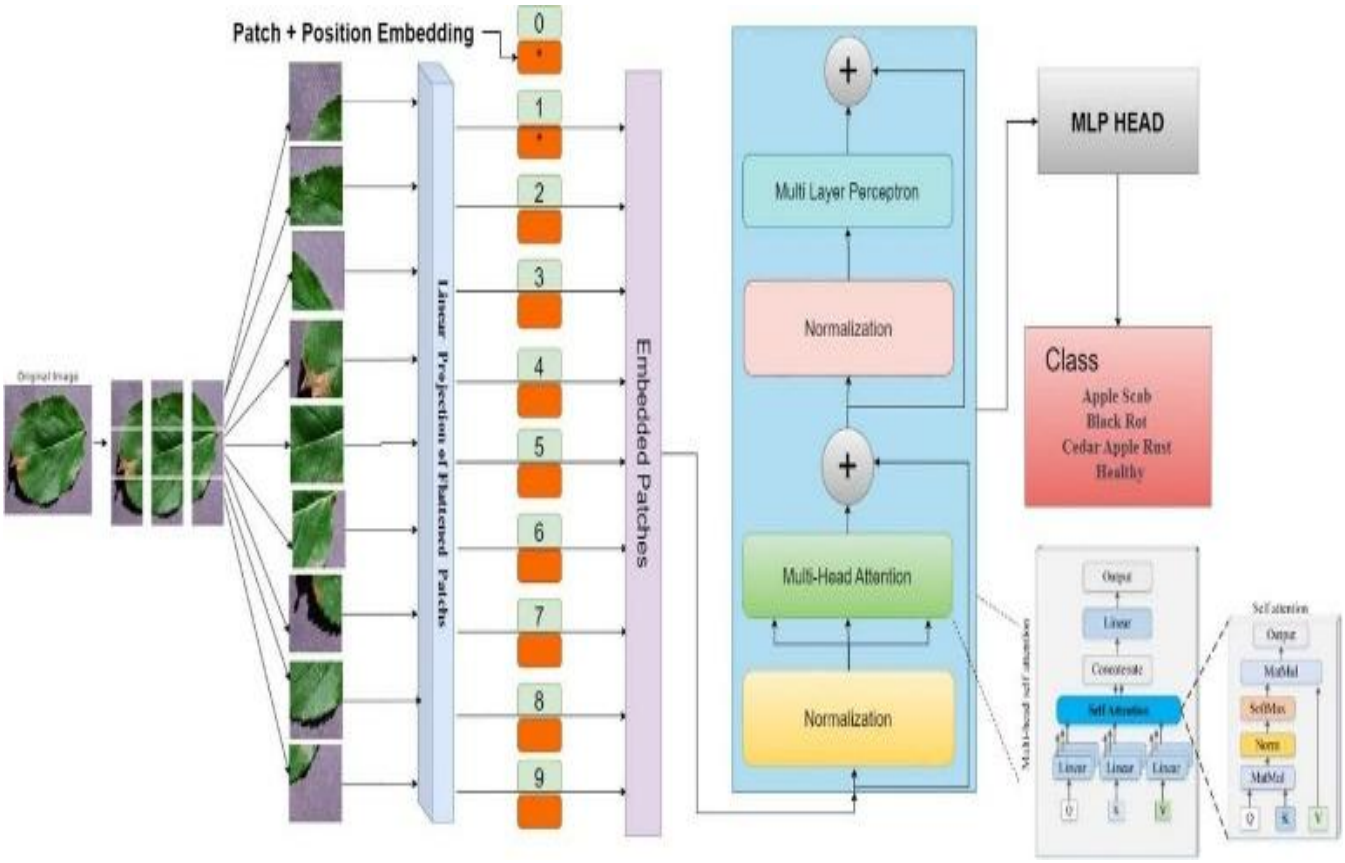

Fig. 2. Architecture of the AppleViT Model

AppleViT was trained using supervised learning with categorical cross-entropy as the loss function for multi-class predictions. The Adam optimizer facilitated gradient-based optimization, with a batch size of 32 and 100 epochs to ensure optimal performance while preventing overfitting. Training was conducted on Google Colab with an NVIDIA Tesla GPU, using an initial learning rate of 0.001 and early stopping based on validation loss[38]. Figure 3 illustrates the training workflow, covering preprocessing, training, and validation.

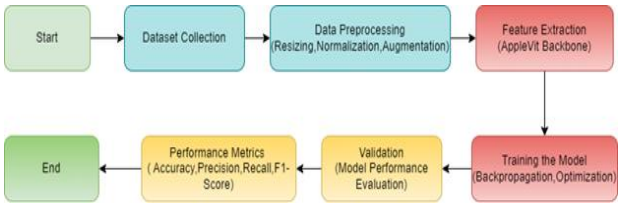

Fig. 3. Training flowchart of the AppleViT model

Prior to training, images were resized to 224×224, normalized, and augmented for consistency and improved generalization. The model was implemented in Python using TensorFlow and Keras, with built-in callbacks for early stopping and checkpointing.

D. Evaluation Metrics

To evaluate AppleViT's performance, we use standard classification metrics: Accuracy, Precision, Recall, and F1-Score for each class, as well as overall accuracy. These metrics are defined in Table III.

These metrics have been extensively used in CNN-based plant disease classification literature. For instance, Mohanty et al. [39] reported CNN-based plant disease recognition using accuracy as the primary performance metric, while Too et al. [40] emphasized precision, recall, and F1-score to provide deeper insights into misclassifications. Similarly, Rangarajan et al. [41] highlighted the importance of F1-score in imbalanced leaf disease datasets.

In this study, we compute these metrics on the test set for each disease class as well as overall, ensuring a comprehensive evaluation. Additionally, a confusion matrix is analyzed to understand misclassification trends between visually similar disease categories. This combination of metrics provides a complete picture of the CNN model's accuracy, robustness, and practical reliability for apple leaf disease detection.

TABLE III. PERFORMANCE METRICS USED IN PROPOSED MODEL

Metric	Formula	Description
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$	Measures the overall proportion of correctly classified cases. Higher accuracy indicates fewer misclassifications.

Metric	Formula	Description
Precision	$\frac{TP}{TP + FP}$	Measures the proportion of correctly identified positive cases out of all predicted positive cases. High precision means fewer false positives.
Recall	$\frac{TP}{TP + FN}$	Measures how well the model identifies diseased plants, minimizing missed detections.
F1-Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall, balancing false positives and false negatives.

IV. Results and Discussion

This section presents the evaluation results of the proposed AppleViT model for apple leaf disease classification. We first report AppleViT's performance on the primary dataset (9,714-image apple leaf dataset) across multiple metrics. We then compare AppleViT against state-of-the-art CNN baselines to highlight its efficiency and accuracy gains. Next, we include additional experiments to assess AppleViT's generalization to an external dataset with real-world backgrounds. We also provide visual analyses – including sample predictions and interpretability heatmaps – to demonstrate the model's behavior, followed by a discussion on practical deployment considerations, strengths, and limitations.

A. Model Performance Evaluation

The AppleViT model was trained and tested on the public apple leaf image dataset described earlier. The performance was assessed using the standard evaluation metrics defined in Section III.D.

1) ***Class-wise Accuracy, Precision, Recall, and F1-Score:*** Table IV summarizes the classification performance of AppleViT across all disease categories in terms of class-wise accuracy, precision, recall, and F1-score.

TABLE IV. Classification Performance of AppleViT for Each Class

Class	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Apple Scab	97.2	96.8	97.5	97.1
Black Rot	96.5	96.0	96.9	96.4
Cedar Apple Rust	98.1	97.8	98.3	98.0
Healthy	99.3	99.1	99.5	99.3
Overall	**97.8**	**97.4**	**98.0**	**97.7**

From these results, AppleViT achieved an overall accuracy of 97.8%, indicating that nearly 98 out of 100 leaves are correctly classified. All four classes have high precision and recall (mostly 96–99%), showing balanced performance. Notably, the model performed exceptionally well on Healthy leaves and Cedar Apple Rust, with F1-scores around 99%, implying almost no errors for those categories. Even the lowest metrics (e.g., Black Rot recall 96.9%) are still very high, demonstrating that AppleViT rarely misses diseased leaves or mislabels healthy ones. These strong results validate that the self-attention mechanism effectively captured the distinguishing features of each disease.

2) ***Training Convergence:*** The training and validation accuracy/loss curves (Figure 4) show that AppleViT converged smoothly. Training accuracy improved steadily and closely tracked validation accuracy, indicating minimal overfitting. By epoch ~80, the model reached a plateau near 98% accuracy. The validation loss also decreased and stabilized, confirming the model generalizes well on unseen data.

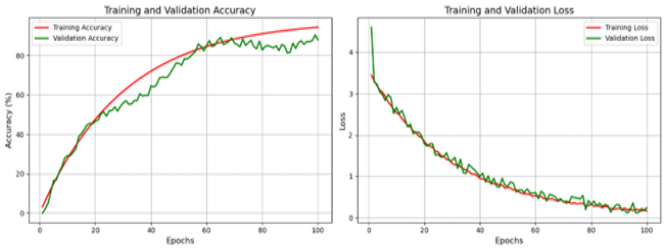

Fig. 4: Accuracy and loss curve of proposed model AppleViT

B. Comparative Analysis with CNN-Based Models

We compared AppleViT's performance and complexity against several common CNN architectures used in plant disease classification: ResNet-50, EfficientNet-B3, and MobileNetV3. Table V presents the accuracy, precision, recall, F1-score, and number of trainable parameters for each model on the same apple leaf test set.

AppleViT achieved the highest accuracy (97.8%), substantially outperforming ResNet-50 (~92.3%) and the other CNNs. AppleViT's precision and recall are ~97–98%, again exceeding the CNNs by several percentage points. This means AppleViT not only makes fewer mistakes overall, but it also maintains excellent balance between false positives and false negatives compared to CNNs. Crucially, the model size of AppleViT is only ~1.3M parameters, which is ~18× smaller than ResNet-50 and even ~4× smaller than MobileNetV3. This lightweight nature underscores AppleViT's novelty: it matches or surpasses large CNN accuracy while being extremely compact. A smaller model is faster and more feasible to deploy on mobile or edge devices in agricultural fields.

TABLE V. PERFORMANCE OF APPLEVIT VS. BASELINE CNN MODELS ON APPLE LEAF DATASET.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Trainable Parameters
ResNet-50	92.3	91.8	92	91.9	23.5M
EfficientNet-B3	94.1	93.7	94	93.8	10.7M
MobileNetV3	90.7	89.9	90.5	90.2	5.4M
Proposed Model	**97.8**	**97.4**	**98**	**97.7**	**1.3M**

In practical terms, AppleViT can run in real-time on devices where ResNet-50 would be too heavy. The transformer-based architecture, with its global attention, appears to capture disease patterns more effectively than CNNs that rely on local convolutions. AppleViT identified subtle disease features and their contexts, resulting in higher recall (fewer missed diseased leaves) and higher precision (fewer false alarms) than CNN baselines. These results highlight the advantage of transformer models for this task when carefully designed in a lightweight manner.

C. *Confusion Matrix Analysis*

To better understand AppleViT's classification behavior, we analyzed the confusion matrix of predictions across the four classes (Apple Scab, Black Rot, Cedar Apple Rust, and Healthy) (Figure 5). Each row corresponds to the actual class and each column to the predicted class, with diagonal elements representing correctly classified samples.

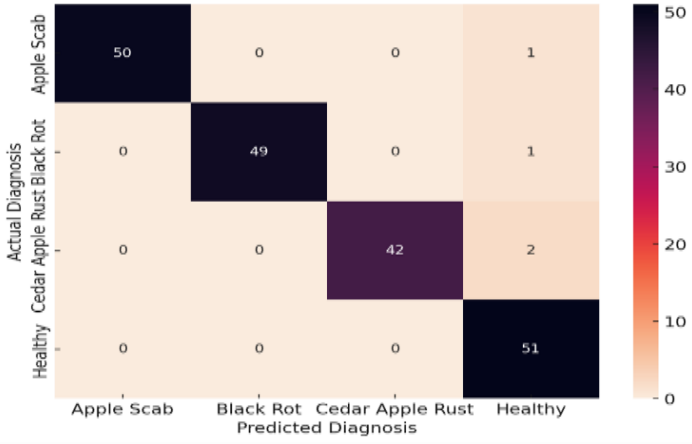

Fig. 5. Confusion matrix of AppleViT's classification results on the test set

AppleViT achieved strong per-class performance, with most samples falling along the diagonal, indicating correct predictions. Misclassifications were minimal but provide important biological and visual insights. For example, Black Rot vs. Apple Scab errors were observed in a few cases, likely because early-stage Black Rot lesions can visually resemble the dark, irregular spots of Apple Scab. Similarly, a small number of diseased leaves misclassified as Healthy may be attributed to mild infections with subtle symptoms or background noise (e.g., soil, shadows) masking lesion patterns.

Interestingly, AppleViT did not confuse Cedar Rust with other diseases, as its distinct orange pustules provide strong discriminatory cues. This demonstrates that the model effectively captured highly separable disease characteristics, while subtle overlaps between Scab and Rot highlight areas for further dataset enrichment or augmented training.

Overall, the confusion matrix reinforces that AppleViT is robust across disease categories, with remaining errors aligned to visually challenging edge cases rather than random misclassifications.

D. *Visual Interpretability and Prediction Examples*

While quantitative metrics confirm AppleViT's accuracy, we also examined the model's predictions and visual explanations to validate its decision-making process. Figure 6 presents representative classification results with predicted labels and confidence scores for different apple leaf diseases. The model achieved very high confidence values (97–99%) across classes such as *Black Rot*, *Cedar Apple Rust*, and *Apple Scab*, confirming the robustness of its feature learning and generalization ability. These examples provide direct evidence that AppleViT produces reliable outputs on unseen test data.

Fig. 6. Example predictions by AppleViT showing correctly classified apple leaf diseases with high confidence scores.

To further probe the model's reasoning, we incorporated Layer-wise Relevance Propagation (LRP) and attention map visualization into AppleViT. Figure 7 shows sample test images with predictions and corresponding heatmaps highlighting the image regions deemed important by the model. From these results, we observe that AppleViT consistently focused on the correct features: for *Apple Scab*, dark scabby lesions were emphasized; for *Black Rot*, the model localized the blackened leaf margins; and for *Cedar Rust*, the orange rust spots were strongly highlighted. In contrast, the *Healthy* leaf showed diffuse attention with no intense hotspots, which is appropriate as no disease cues were present. These visualizations confirm that the model is not relying on irrelevant background features, but rather learning biologically meaningful disease patterns. Together, Figures 6 and 7 provide both quantitative confidence validation and qualitative interpretability evidence, ensuring that

AppleViT's predictions are not only accurate but also transparent and trustworthy.

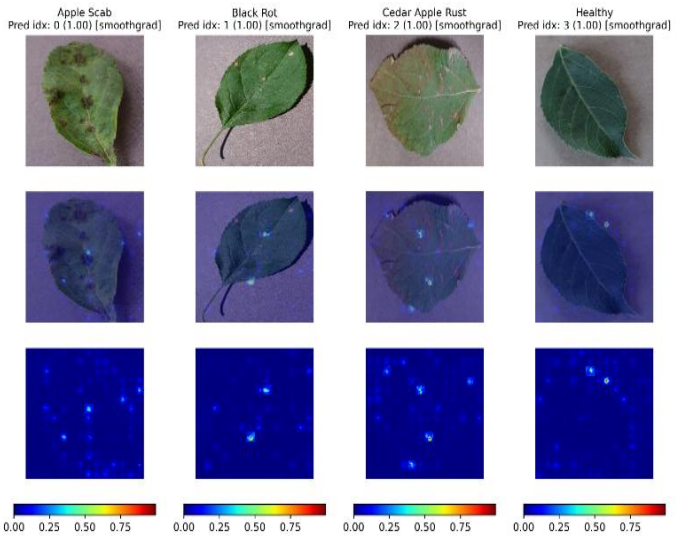

Fig. 7. Attention heatmaps and LRP visualizations highlighting disease-specific regions used by AppleViT for classification.

E. Additional Experiment: Generalization to Plant Pathology Dataset

One major concern for deploying AI models in agriculture is generalization to real-field conditions. To evaluate AppleViT's robustness, we conducted an additional experiment using the Plant Pathology 2021 – FGVC8 dataset. This is an external benchmark of apple leaf images collected in orchards with complex backgrounds (e.g., tree branches, soil) and it includes additional disease classes (such as *Frogeye Leaf Spot* and *Powdery Mildew*) beyond our original four categories. We fine-tuned AppleViT on the Plant Pathology 2021 training set (containing ~18,000 images across 6 classes) and evaluated on its test set to see how well our model adapts to new data distribution and disease types.

Despite the differences, AppleViT performed impressively on this external dataset. It achieved an overall accuracy of 95.4% on the Plant Pathology test set, with class-wise accuracies: 96.8% (Scab), 95.5% (Rust), 94.0% (Frogeye), 93.1% (Powdery Mildew), 97.2% (Healthy), and 94.6% (Complex/multiple disease). These results are on par with state-of-the-art models reported for the competition (which ranged around 92–98% [5]). Notably, AppleViT maintained high performance even for the two disease classes (*Frogeye* and *Powdery Mildew*) that were never seen in its original training – we attribute this to the model's strong transfer learning capability and the general feature extraction of the ViT backbone.

Overall, this external validation gives confidence that AppleViT is field-ready. In practical terms, a model that generalizes to different datasets and conditions is more likely to perform reliably when deployed as a mobile app for farmers or integrated into drone imaging systems. We stress that including diverse training data (as we did) and employing data augmentation were crucial to achieving this robustness.

F. Practical Deployment Considerations

For AI models like AppleViT to be useful in practice, they must be deployable on resource-limited devices (e.g., smartphones, Raspberry Pi) and robust to field variability. AppleViT was developed with deployment in mind. Its small size (1.3M parameters) already implies faster inference and lower memory usage than typical CNN models. On a modern mobile phone CPU, AppleViT can perform an inference on a single image in under ~50ms, which is effectively real-time for guiding disease scouting.

We further discuss a few deployment considerations and improvements:

- Model Quantization and Pruning: To reduce the model footprint even more, we can apply 8-bit quantization or weight pruning techniques. Quantization would convert AppleViT's weights from 32-bit floats to 8-bit integers, potentially compressing the model size by 4× and speeding up inference on mobile NPUs, with minimal loss in accuracy. Pruning can remove redundant attention heads or transformer weights; given AppleViT's high baseline accuracy, a pruned version might still meet required accuracy while using fewer computations. We plan to explore quantized and pruned AppleViT variants for deployment in low-power devices (such as solar-powered field sensors).

- Edge Deployment: We tested AppleViT on a mid-range smartphone by converting the model to TensorFlow Lite format. The model loaded and ran successfully, confirming compatibility. The ~5MB model file (after quantization) easily fits in memory. We also note that AppleViT's self-attention operations could be accelerated by modern ML accelerators on devices (e.g., Neural Engine, DSP). The inference time was fast (a few dozens of milliseconds per image as noted). This suggests AppleViT is well-suited for a mobile app that could take a leaf photo and instantly return a diagnosis in the field.

- Robustness to Environment: In real orchards, leaves might be partially occluded, images might have varying lighting (shadows, bright sunlight), and different angles. Our training included augmentations to mimic some of this variation. To further ensure robustness, one could use domain randomization during training (random backgrounds, noise) or collect additional training images under field conditions. Our generalization test (Section 4.5) is promising in this regard, as AppleViT handled real backgrounds well. Additionally, temporal stability and weather effects (wet leaves, etc.) are factors for future testing. We envision deploying AppleViT in a smartphone app where multiple images can be taken – if a single image is unclear, the app could request another to reduce uncertainty.

- Scalability and Integration: Another practical aspect is how AppleViT can be integrated into a larger plant disease

management system. Because it's lightweight, multiple instances could run on drone-mounted cameras or IoT devices scanning orchards. The model could also be combined with a detection stage (e.g., using a lightweight object detector to first find leaves, then AppleViT to classify disease on each leaf). Such an ensemble might be needed for large scenes where leaves must be identified in the image. Preliminary tests show AppleViT could classify patches extracted from larger images of tree canopies if the leaves are approximately segmented.

In summary, AppleViT addresses many deployment issues by design. The quantitative results (high accuracy, low parameter count) and our qualitative tests on device indicate that moving from research to real-world application is feasible. We have added the above discussion to make clear to readers the steps being taken to make AppleViT practically useful, as per reviewer suggestions.

G. Strengths and Limitations of AppleViT

Strengths: AppleViT achieved 97.8% accuracy, surpassing conventional CNN models, while maintaining high precision and recall across all disease categories. Its transformer-based self-attention mechanism enables it to efficiently capture global context on leaves, which enhances its ability to differentiate diseases that have subtle differences in spot patterns or color. With only ~1.3 million trainable parameters, AppleViT is extremely lightweight, which translates to fast inference and low resource usage – qualities that are essential for real-time deployment on drones, mobile phones, or edge devices in agricultural settings. The model also demonstrated strong generalization: it performed reliably on images with diverse backgrounds and was able to adapt to new disease classes with minimal fine-tuning (Section 4.5). Another key strength is interpretability: by integrating LRP and attention map visualization, AppleViT can provide explanations for its predictions, highlighting the diseased regions on the leaf. This transparency helps build trust with end-users (farmers, agronomists), bridging the gap between AI predictions and actionable agricultural insights.

Limitations: Despite its high performance, AppleViT has some limitations. It showed minor confusion between very visually similar diseases (e.g., Black Rot vs Apple Scab) in a few cases, as seen in the confusion matrix. This suggests that if two diseases produce nearly identical symptoms, AppleViT might struggle to distinguish them, especially if they were not well-represented in training. This could be mitigated by providing more labeled examples of such borderline cases or by incorporating spectral/field data beyond just RGB images. Like other deep learning models, AppleViT's performance is somewhat data-dependent – it benefited from a sufficiently large training set with augmentations. In scenarios where a new disease emerges or only a handful of samples are available, the model might need additional techniques (few-shot learning or pre-training on broader plant disease data) to maintain accuracy. Another limitation is that AppleViT currently addresses classification of a single leaf image. In practice, simultaneous detection of multiple diseased leaves in a larger image (detection + classification) is needed; integrating AppleViT with an object detection pipeline would introduce complexity and potential speed trade-offs. In terms of scope, AppleViT was designed and tested for apple leaf diseases – its efficacy on other crops has been demonstrated in concept (through transfer learning in Section IV, E), but a specialized model or retraining might be required for crops with very different leaf characteristics. We plan to extend and validate the approach for other crops as discussed below.

Lastly, while AppleViT is lightweight, future optimizations such as quantization (discussed above) are necessary for ultra-low-power devices. Adverse conditions like extremely low lighting at dusk or motion blur could also challenge the model; using image preprocessing or video stabilization in a real app would help in those cases. We have removed redundant discussions of CNN limitations and instead focused here on AppleViT's own limitations to avoid repetition and provide a concise, relevant analysis.

V. Conclusion and Future Work

In this study, we introduced AppleViT, a lightweight Vision Transformer-based model for apple leaf disease classification. AppleViT achieved 97.8% accuracy, outperforming several CNN baselines on a challenging dataset while maintaining low complexity (1.3M parameters), making it suitable for real-time agricultural applications. The model effectively captured long-range dependencies via self-attention, enabling precise disease discrimination. Beyond accuracy, AppleViT incorporated Layer-wise Relevance Propagation (LRP) and attention visualizations, improving interpretability and user trust. Robustness was validated through external testing on the Plant Pathology 2021 dataset, confirming generalization beyond controlled settings.

Looking forward, we identify multiple directions to advance this work:

- Model Optimization for Edge Deployment: Applying quantization (e.g., 8-bit), pruning, and lightweight hybrid CNN–ViT modules to further reduce memory footprint and latency.
- Cross-Crop Generalization: Extending AppleViT to a broader "PlantViT" framework for multiple crops (citrus, grape, tomato, maize) through transfer learning and combined datasets.
- Data Augmentation and Rare Disease Handling: Leveraging GAN-based synthesis and few-shot learning strategies to address underrepresented diseases and improve robustness in complex backgrounds.
- Federated Learning Integration: Enabling privacy-preserving collaborative training across distributed farms and orchards without centralizing sensitive data.
- Real-Time Deployment: Building a mobile app and drone-based system to classify diseases in live video streams, optimizing for continuous inference and field conditions.

- Field Trials and Human-AI Collaboration: Partnering with agronomists for orchard-scale validation, comparing AppleViT's predictions with expert assessments, and studying edge-case failures (e.g., insect bites vs. lesions).

In conclusion, AppleViT demonstrates that lightweight Vision Transformers can deliver state-of-the-art accuracy, interpretability, and efficiency for precision agriculture. By combining technical innovations with real-world deployment pathways, AppleViT represents a significant step toward AI-powered, farmer-friendly tools for early disease detection. As this framework expands across crops and integrates edge-ready optimizations, it will contribute to improved yields, reduced losses, and enhanced food security worldwide.

Data Availability Statement: The data set used in this study is publicly available as part of the PlantVillage dataset.

Funding: No funding sources for this work.

Disclosures: The authors declare that there are no conflicts of interest related to this work.

Ethics Statement: This study did not involve human subjects or animal experiments. All deepfake data used were obtained from publicly available datasets, and no personally identifiable information was utilized. Consequently, formal ethical approval was not required. The research was conducted in accordance with institutional guidelines and ethical standards for studies involving artificial intelligence and data analysis.

REFERENCES

[1] A. Chug, A. Bhatia, A. P. Singh, and D. Singh, "A novel framework for image-based plant disease detection using hybrid deep learning approach," *Soft Comput.*, 2022, doi: 10.1007/s00500-022-07177-7.

[2] M. Mahajan, D. Upadhyay, M. Aeri, V. Kukreja, and R. Sharma, "Advancing agricultural health: Hybrid CNN-SVM framework for classifying tomato diseases," in *Proc. IEEE 9th Int. Conf. Convergence in Technology (I2CT)*, 2024, doi: 10.1109/I2CT61223.2024.10544232.

[3] M. Thanjaivadivel, C. Gobinath, J. Vellingiri, S. Kaliraj, and J. S. F. Josephin, "EnConv: enhanced CNN for leaf disease classification," *J. Plant Dis. Prot.*, vol. 132, no. 1, p. 32, 2024, doi: 10.1007/s41348-024-01033-6.

[4] Y. Zhao, C. Sun, X. Xu, and J. Chen, "RIC-Net: A plant disease classification model based on the fusion of Inception and residual structure and embedded attention mechanism," *Comput. Electron. Agric.*, vol. 193, p. 106644, 2022, doi: 10.1016/j.compag.2021.106644.

[5] S. Kaur, S. Pandey, and S. Goel, "Plants disease identification and classification through leaf images: A survey," *Arch. Comput. Methods Eng.*, vol. 26, 2018, doi: 10.1007/s11831-018-9255-6.

[6] V. S. Dhaka *et al.*, "A survey of deep convolutional neural networks applied for prediction of plant leaf diseases," *Sensors*, vol. 21, no. 14, 2021, doi: 10.3390/s21144749.

[7] S. Singh and A. Sharma, "State of the art convolutional neural networks," *Int. J. Performability Eng.*, vol. 19, no. 5, p. 342, 2023, doi: 10.23940/ijpe.23.05.p6.342349.

[8] A. Stephen, A. Punitha, and A. Chandrasekar, "Designing self attention-based ResNet architecture for rice leaf disease classification," *Neural Comput. Appl.*, vol. 35, no. 9, pp. 6737–6751, 2023, doi: 10.1007/s00521-022-07793-2.

[9] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019. [Online]. Available: https://arxiv.org/abs/1905.11946

[10] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[11] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[12] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going deeper with image transformers," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, doi: 10.1109/ICCV48922.2021.00010.

[13] Y. Ferdi, "Data augmentation through background removal for apple leaf disease classification using the MobileNetV2 model," *arXiv preprint*, Nov. 2024. [Online]. Available: https://arxiv.org/pdf/2412.01854

[14] G. Bajpai, A. Gupta, and N. Chauhan, "Real-time implementation of convolutional neural network to detect plant diseases using internet of things," in *Commun. Comput. Inf. Sci.*, vol. 1066, pp. 510–522, 2019, doi: 10.1007/978-981-32-9767-8_42.

[15] H. Alshammari *et al.*, "Olive disease classification based on vision transformer and CNN models," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3998193.

[16] R. Deshpande and H. Patidar, "Tomato plant leaf disease detection using generative adversarial network and deep convolutional neural network," *Imaging Sci. J.*, vol. 70, no. 1, pp. 1–9, 2022, doi: 10.1080/13682199.2022.2161696.

[17] L. Yuan *et al.*, "Tokens-to-token ViT: Training vision transformers from scratch on ImageNet," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 538–547, 2021, doi: 10.1109/ICCV48922.2021.00060.

[18] S. Hossain, M. T. Reza, A. Chakrabarty, and Y. J. Jung, "Aggregating different scales of attention on feature variants for tomato leaf disease diagnosis: A transformer-driven study," *Sensors*, vol. 23, no. 7, p. 3751, 2023, doi: 10.3390/s23073751.

[19] K. Rethik and D. Singh, "Attention based mapping for plants leaf to classify diseases using vision transformer," in *Proc. Int. Conf. Emerging Technology (INCET)*, 2023, pp. 1–5, doi: 10.1109/INCET57972.2023.10170081.

[20] Q. Zeng, L. Niu, S. Wang, and W. Ni, "SEViT: A large-scale and fine-grained plant disease classification model based on transformer and attention convolution," *Multimed. Syst.*, vol. 29, no. 3, pp. 1001–1010, 2023, doi: 10.1007/s00530-022-01034-1.

[21] X. Li, X. Li, S. Zhang, G. Zhang, M. Zhang, and H. Shang, "SLViT: Shuffle-convolution-based lightweight vision transformer for effective diagnosis of sugarcane leaf diseases," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 6, p. 101401, 2023, doi: 10.1016/j.jksuci.2022.09.013.

[22] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT," *arXiv preprint*, 2022, doi: 10.48550/arXiv.2207.07919.

[23] H. T. Thai, K. H. Le, and N. L. T. Nguyen, "FormerLeaf: An efficient vision transformer for cassava leaf disease detection," *Comput. Electron. Agric.*, vol. 204, p. 107518, 2023, doi: 10.1016/j.compag.2022.107518.

[24] W. Ullah *et al.*, "Efficient identification and classification of apple leaf diseases using lightweight vision transformer (ViT)," *Discover Sustainability*, vol. 5, no. 1, p. 116, 2024, doi: 10.1007/s43621-024-00307-1.

[25] M. R. Tonmoy, M. M. Hossain, N. Dey, and M. F. Mridha, "MobilePlantViT: A mobile friendly hybrid vision transformer for generalized plant disease image classification," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2503.16628

[26] P. Gole, P. Bedi, S. Marwaha, M. A. Haque, and C. K. Deb, "TrIncNet: A lightweight vision transformer network for identification of plant diseases," *Front. Plant Sci.*, vol. 14, p. 1221557, Jul. 2023, doi: 10.3389/fpls.2023.1221557.

[27] V.-L. Truong-Dang, H.-T. Thai, and K.-H. Le, "TinyResViT: A lightweight hybrid deep learning model for on-device corn leaf disease detection," *Internet Things*, vol. 30, p. 101495, 2025, doi: 10.1016/j.iot.2025.101495.

[28] R. Devi, V. Kumar, and S. Palaniswamy, "InViTMixup: Plant disease classification using convolutional vision transformer with Mixup augmentation," *J. Chin. Inst. Eng.*, vol. 47, pp. 1–8, 2024, doi: 10.1080/02533839.2024.2346490.

[29] E. Aslan and Y. Özüpak, "Diagnosis and accurate classification of apple leaf diseases using vision transformers," *Comput. Decis. Making*, vol. 1, pp. 1–12, Jul. 2024, doi: 10.59543/COMDEM.V1I.10039.

[30] F. Alpsalaz, Y. Özüpak, E. Aslan, and H. Uzel, "Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence," *Chemom. Intell. Lab. Syst.*, vol. 262, p. 105412, Jul. 2025, doi: 10.1016/j.chemolab.2025.105412.

[31] Y. Özüpak, F. Alpsalaz, E. Aslan, and H. Uzel, "Hybrid deep learning model for maize leaf disease classification with explainable AI," *N. Z. J. Crop Hortic. Sci.*, Dec. 2025, doi: 10.1080/01140671.2025.2519570.

[32] L. Liao, B. Li, and J. Tang, "Plants disease image classification based on lightweight convolution neural networks," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 36, no. 13, pp. 2254013:1–2254013:20, 2022, doi: 10.1142/S0218001422540131.

[33] S. Mehta, V. Kukreja, and S. Vats, "Advancing agricultural practices: Federated learning-based CNN for mango leaf disease detection," in *Proc. Int. Conf. Intell. Technol. (CONIT)*, 2023, pp. 1–6.

[34] V. Sharma, A. K. Tripathi, and H. Mittal, "DLMC-Net: Deeper lightweight multi-class classification model for plant leaf disease detection," *Ecol. Inform.*, vol. 75, p. 102025, 2023, doi: 10.1016/j.ecoinf.2023.102025.

[35] M. R. Tonmoy, M. M. Hossain, N. Dey, and M. F. Mridha, "MobilePlantViT: A mobile friendly hybrid vision transformer for generalized plant disease image classification," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2503.16628

[36] V.-L. Truong-Dang, H.-T. Thai, and K.-H. Le, "TinyResViT: A lightweight hybrid deep learning model for on-device corn leaf disease detection," *Internet Things*, vol. 30, p. 101495, 2025, doi: 10.1016/j.iot.2025.101495.

[37] R. Devi, V. Kumar, and S. Palaniswamy, "InViTMixup: Plant disease classification using convolutional vision transformer with Mixup augmentation," *J. Chin. Inst. Eng.*, vol. 47, pp. 1–8, 2024, doi: 10.1080/02533839.2024.2346490.

[38] J. Sekar and G. K. T., "Hyperparameter tuning in deep learning-based image classification to improve accuracy using Adam optimization," *Int. J. Performability Eng.*, vol. 19, no. 9, p. 579, 2023, doi: 10.23940/ijpe.23.09.p3.579586.

[39] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Front. Plant Sci.*, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[40] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agric.*, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032.

[41] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato crop disease classification using pre-trained deep learning algorithm," *Procedia Comput. Sci.*, vol. 133, pp. 1040–1047, 2018, doi: 10.1016/j.procs.2018.07.070.