CHAPTER-1

INTRODUCTION

1.1 Project Overview

Plant diseases threaten global agriculture, causing economic losses and food insecurity. The FAO reports that nearly 40% of crop yield is lost annually to plant diseases. Timely, accurate disease identification is vital for sustainable agriculture and food security.

Recent progress in deep learning has made it possible to automatically recognize plant diseases from images. Convolutional Neural Networks (CNNs) have been successful, but they mainly extract local features and often miss broader patterns in plant images. Vision Transformers (ViTs) address this by capturing global spatial relationships, which helps improve classification accuracy.

Traditional ViT models are effective but require extensive computational and memory resources. This makes them difficult to deploy in resource-constrained settings like mobile or edge devices. To address these issues, this work proposes **MobilePlantViT**. It is a lightweight hybrid Vision Transformer architecture optimized for efficient plant disease classification.

Our model uses **Ghost Convolutions** for efficient feature extraction, **Coordinate Attention** to focus on important disease areas, and **Fused-Inverted Residual** Blocks to improve how features are represented. We also use **Linear Differential Attention** to lower computational demands and a **Bottleneck Feed Forward Network (FFN)** to keep the model small without losing accuracy. Together, these features help MobilePlantViT balance accuracy, speed, and efficiency, making it practical for use on low-power devices in real-world agriculture.

1.2 Background and Motivation

The identification of plant diseases has traditionally relied on manual inspection by agricultural experts, a process that is time-consuming, subjective, and often inaccurate under field conditions. The introduction of artificial intelligence and computer vision has revolutionized this domain by enabling rapid and objective disease recognition. However, most state-of-the-art deep learning models remain computationally intensive, limiting their practicality for on-site diagnosis, particularly in rural regions where computational resources are limited.

Vision Transformers have demonstrated exceptional capabilities in visual understanding by leveraging self-attention mechanisms. Nevertheless, their high computational overhead poses a barrier to real-time applications. There is thus a growing need for lightweight ViT architectures that retain the accuracy of transformer-based models while minimizing resource consumption.

This research is motivated by the necessity to design an optimized Vision Transformer that maintains high accuracy in disease detection while remaining computationally efficient. **MobilePlantViT** addresses this challenge through architectural simplifications, hybrid feature learning, and parameter reduction, thereby enabling the deployment of AI-driven plant disease recognition systems in practical agricultural settings.

1.3 Objectives

The objectives of this research are as follows:

1. To design a lightweight hybrid Vision Transformer architecture for efficient plant disease classification.
2. To minimize computational complexity and model parameters while maintaining high classification accuracy.
3. To integrate efficient modules such as Ghost Convolutions, Coordinate Attention, Fused-Inverted Residual Blocks, Linear Differential Attention, and Bottleneck FFN.
4. To improve feature representation and discrimination for diverse plant disease datasets.
5. To ensure cross-dataset generalization and adaptability for real-world deployment.
6. To achieve compatibility with low-power or edge devices for real-time inference.
7. To evaluate the model using standard metrics and benchmark it against existing lightweight ViT architectures.

1.4 Methodology

The methodology adopted for developing MobilePlantViT involves the following stages:

1.4.1 Problem Identification

A detailed analysis of existing deep learning-based approaches for plant disease detection was performed to identify limitations in computational efficiency and scalability.

1.4.2 Dataset Preparation

Datasets were collected from publicly available sources such as **PlantVillage** and other agricultural repositories. Preprocessing techniques, including image resizing, normalization, and augmentation, were applied to ensure uniformity and improve generalization.

1.4.3 Model Architecture Design

The proposed architecture integrates Ghost Convolutions for lightweight feature extraction, Coordinate Attention for spatial emphasis, and Fused-Inverted Residual Blocks for efficient feature transformation. Linear Differential Attention simplifies the attention computation, while a Bottleneck FFN reduces parameter count and enhances inference speed.

1.4.4 Model Training and Evaluation

The model was trained using the Adam optimizer and cross-entropy loss function. Performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Comparisons were made with existing lightweight architectures, including MobileViTv1 and MobileViTv2.

1.4.5 Deployment Optimization

The trained model was optimized for real-time applications using quantization and pruning techniques, ensuring compatibility with mobile and edge hardware through frameworks such as TensorFlow Lite and PyTorch Mobile.

1.5 Potential Impact

The proposed MobilePlantViT architecture contributes to the advancement of smart and sustainable agriculture in several ways:

1. **Enhanced Agricultural Productivity:** Enables early and accurate detection of plant diseases, improving crop yield and management efficiency.
2. **Accessibility and Affordability:** Facilitates the deployment of AI models on low-cost devices, benefiting farmers in resource-limited regions.
3. **Energy Efficiency:** Reduces computational overhead, thereby lowering energy consumption and supporting eco-friendly AI solutions.
4. **Scalability:** The model's modular design allows adaptation for other agricultural tasks such as pest detection and crop health monitoring.
5. **Research Contribution:** Provides a novel and efficient hybrid Vision Transformer framework that can serve as a foundation for future work in lightweight AI models.